

This page Is Inserted by IFW Operations
And is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM_protein - nucleic search, using frame_p2n model

Run on: November 25, 2003, 10:20:54 / Search time 93 Seconds

(without alignments)
2444.218 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTTBEXKPFKLSFLWPIH.....AQSALKEQVAVQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Ygapext 0.5
Xgapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09869433/runat_24112003_171459_18152/app_query.fasta_1.711
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPTC=0
-LOOPEXT=0 -UNITS=bases -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdt
-LIST=45 -DOCLING=200 -THR SCORE=pct -THR MIN=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=p2n -NDBM=ext HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09869433 @CGN 1.1 56 @runat_24112003_171459_18152 -NCPU=6 -ICPU=3
-NO MMAP -LARBQUERRY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCtus.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2617	99.5	1230025	4 US-09-198-452A-1	Sequence 1, Appl
2	1025.5	39.0	1230025	4 US-09-198-452A-1	Sequence 5369, Ap
3	242	9.2	280	4 US-09-313-294A-5369	Sequence 3344, Ap
4	168	6.4	265	4 US-09-313-294A-5369	Sequence 15706, A
5	135	5.1	1365	4 US-09-252-991A-15706	Sequence 15648, A
6	135	5.1	1638	4 US-09-252-991A-15648	Sequence 15612, A
7	135	5.1	2859	4 US-09-252-991A-15612	Sequence 1, Appl
8	133	5.1	1830121	4 US-09-557-884-1	Sequence 1, Appl
9	133	5.1	1830121	4 US-09-643-990A-1	Sequence 1331, Ap
10	122.5	4.7	1467	4 US-09-252-991A-1331	Sequence 1, Appl
11	122.5	4.7	68750	3 US-09-335-409-1	Sequence 1, Appl
12	122.5	4.7	68750	4 US-09-568-102-1	Sequence 1, Appl

13	122.5	4.7	68750	4 US-09-567-969-1	Sequence 1, Appl
14	122.5	4.7	68750	4 US-09-568-480-1	Sequence 1, Appl
15	122.5	4.7	68750	4 US-09-568-486-1	Sequence 1, Appl
16	122.5	4.7	68750	4 US-09-568-472-1	Sequence 1, Appl
17	122.5	4.7	68750	4 US-09-567-899-1	Sequence 1, Appl
18	120.5	4.6	1605	4 US-09-252-991A-14154	Sequence 14154, A
19	120.5	4.6	1638	4 US-09-252-991A-16400	Sequence 16400, A
20	120.5	4.6	1824	4 US-09-252-991A-14743	Sequence 14243, A
21	120.5	4.6	2228	4 US-09-252-991A-16088	Sequence 16088, A
22	120.5	4.6	2489	4 US-09-252-991A-16193	Sequence 16193, A
23	119.5	4.5	2403	4 US-09-134-001C-2818	Sequence 2818, Ap
24	119	4.5	2823	4 US-09-328-352-4039	Sequence 4039, Ap
25	118	4.5	15363	4 US-08-961-527-139	Sequence 139, Ap
26	117.5	4.5	77536	4 US-09-410-551B-1	Sequence 1, Appl
27	116	4.4	1436	4 US-08-961-527-296	Sequence 296, Ap
28	115.5	4.4	1227	4 US-09-107-532A-2909	Sequence 2909, Ap
29	115.5	4.4	1641	4 US-09-252-991A-533	Sequence 533, Ap
30	114.5	4.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl
31	114.5	4.4	4411529	3 US-09-103-840A-1	Sequence 1, Appl
32	114	4.3	1245	4 US-09-328-352-2835	Sequence 2835, Ap
33	113.5	4.3	1647	4 US-09-328-352-2479	Sequence 2479, Ap
34	113.5	4.3	2431	4 US-09-221-017B-351	Sequence 351, Ap
35	113.5	4.3	6755	3 US-08-931-999-4	Sequence 4, Appl
36	113.5	4.3	1830121	4 US-09-557-884-1	Sequence 1, Appl
37	113.5	4.3	1830121	4 US-09-643-990A-1	Sequence 1800, Ap
38	113	4.3	1437	4 US-09-134-001C-1800	Sequence 1868, Ap
39	113	4.3	1500	4 US-09-252-991A-1868	Sequence 1717, Ap
40	113	4.3	1737	4 US-09-252-991A-1717	Sequence 28, Appl
41	113	4.3	3848	3 US-09-112-096-28	Sequence 14, Appl
42	113	4.3	5668	3 US-09-112-096-14	Sequence 1, Appl
43	113	4.3	4411529	3 US-09-103-840A-1	Sequence 848, Ap
44	112.5	4.3	1260	4 US-09-328-352-848	Sequence 2007, Ap
45	112	4.3	1767	4 US-09-252-991A-2007	

ALIGNMENTS

RESULT 1
US-09-198-452A-1
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OR INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURES:
NAME/KEY: misc feature
LOCATION: (1)..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t

```
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc.feature
LOCATION: (810001)..(825000)
```

Db 404970 CGTGCATTGTTGGGCTTCAAAGTTGAGAGCTTCGGTTCTTGAAGGTGTAAGATCCCTTGG 405029

Qy 221 GYIIESEIELEUARGLEUEMECALMETTHRIIEVALSERGIYLEUVALLEUMECALA 240

Db 405030 GGAATTTCTTAAAGCTCTTTTGAATGGCTATGACTATTGATCTGAGCACTTGTTCTTATGGCC 405089

Qy 241 SETTYTTRTPPIIEANLYASERVALLEUTHIRASPROARGHETRYASNPGLUGLU 260

Db 405090 AGTACGTCGGAGCAATAAGAACGTATTGACCATCTCGCTCTTAATCCAGAAAGA 405149

Qy 261 METGINTYEGGIYLSYSGIYALALYEPOLYSMEASMETLYSASPSERPHILEUTYR 280

Db 405150 ATGCAAAAGGGGAAAAAGAGTGTCTTAACCTTAATAATGATGAAAGATAGCTTCTCTAT 405209

Qy 281 LEUASPARSERPROTYRIIELEULEULEUTHIRLEUENVALIEALATYGIYILECYS 300

Db 405210 CTTCGTAAGTCTCTTAATATTCTTTATTTAGCTCTCTGGTATTGGCTATGATATTGGC 405269

Qy 301 ILEASLEUIIEGLIUALTHITRPLYSERGINLEULYSLEUGINTYRPROASMETASN 320

Db 405270 ATTAACTTAATCGAAGAGCACTGGAAAAAGCAGCTGAACCTGCATTCATTAATGAAT 405329

Qy 321 ASPTYSERGIUPMETGIYASNPHERPHERTPHRRGIYVALSERVALLEUILE 340

Db 405330 GACTATAGTAGTTCATGAGGGAACCTTCCTCTTGAGCTGGCGTAGATCCGATCTATC 405389

Qy 341 METLEUPHEVALGIYGIYASENVALIEARGLYSPHEGIYTRPLEUTHRGYALALEUVAL 360

Db 405390 ATGCTATTGTGGTGGTAAACGTATTCGTAAATTTGGATGGTTAACTGAGCCCTTAGTC 405449

Qy 361 THIRPROVALMETVALLEULEUTHRGYILEVALPHEPHEALAEUVALIEPHEARGASN 380

Db 405450 ACTCCTCATAGTTCCTCTCAACAGATGATGTTTCTTGCTCTTGATCTTATAGAAAC 405509

Qy 381 GIALIASERGIYLEUVALIAMEPHEGIYTHIRTHIRPROLEUMETLEUALAVALIYAL 400

Db 405510 CAACCTTCCTGGCGTGGCTGAGCTATGTTGTAACAACCTCTCAAGCTAGCTGTGGTGTGTC 405569

Qy 401 GYIYALIEGIANUILIELEUSERLYSERTHIRLYSTRYALALEUPHEASPSERTHYLS 420

Db 405570 GGAAGCTATACAGATATTCTTTGGAAATCCACAAAATACGCTCTTTGACTCAACTAAA 405629

Qy 421 GLUMERALATYRIIEPROLEUASPGINGLUGLIMYSVALIYSGIYLSALALALIASP 440

Db 405630 GAATGGCTATATCCCTCTTGACCAAGACAAAAGATCAAAAGSTAAAGCGCTATTGAT 405689

Qy 441 VALVALIALAALARGPHEGIYLYSERGIYGIYALALEUIIEGINGINGIYLEUVAL 460

Db 405690 GTAGTTCCCGCCCGCTTCGGAATAATCAGAGGAGCTTAATCCAAAGGTTTGCTGTT 405749

Qy 461 IIECYGIESERTIEGIALIAMECTHPTHYRLEUVALIIELEULEUPHEIIELE 480

Db 405750 ATCTGGAAGATATGAGCATATGACCCCTTATCTTGCAAGATCTCTTTTATCATAT 405809

Qy 481 ALALIEITRPLEUVALSERALATHRILYSLEUASMLYSLEUPHEUVALAGINSERIALEU 500

Db 405810 GCATTTTGGTTGGTTTTCGCACTAAGTTAAACAAATATTCTTAAAGCGAGTGGCTCTT 405869

Qy 501 LYSGLINGIULVALIAGINGIUSPSERIAPROLASERSER 515

Db 405870 AAAAGAACAAAGAGTGCTCAAGAAATTCAGTCTCTCTCTTCA 405914

RESULT 2

Sequence 1, Application US/09198452A

Patent No. 6559294

GENERAL INFORMATION:

APPLICANT: Grifflais, R.

TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention

TITLE OF INVENTION: and treatment of infection

FILE REFERENCE: 9710-003-999

CURRENT APPLICATION NUMBER: US/09/198,452A

CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (90001)..(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (105001)..(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (120001)..(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (135001)..(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (150001)..(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (165001)..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (180001)..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (195001)..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (210001)..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (240001)..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t

NAME/KEY: misc feature
LOCATION: (330001)..(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

```

LOCATION: (630001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature

```

Alignment Scores:

```

Pred. No.: 6.68e-108 Length: 1230025
Score: 1025.50 Matches: 214
Percent Similarity: 59.12% Conservative: 107
Best Local Similarity: 39.41% Mismatches: 185
Query Match: 38.99% Indels: 37
DB: 4 Gaps: 9

```

US-09-869-433-2 (1-515) x US-09-198-452A-1 (1-1230025)

```

Qy 1 MetThrIysThrGluGluYsPProPhgElyLysLeuArgSerPheLeuTTPProIleHis 20
Db 717505 ATGAGCATCAGCAAGAAAGCAACCTTTTCAAGCGCTGGGCATATCTTTGCTATTTAT 717446
Qy 21 ThrHisGluLeuLysLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 717445 AATTCAGATTTTCTTAAGTTTCTTCACTATTTCTACTAGCGTTTTCGTTGGCTTTTAT 717386
Qy 41 TyrThrValLeuArgAspThrLysAspThrLeuLeuValGlyAlaProGlySerGlyAla 60
Db 717385 TACTGCGCTGCAAAACATGAAGATACCTGCTGATTCGAGTTCAGATGCTGGGCA 717326
Qy 61 GluAlaIleProPheIleLysPheTyrLeuValAlaProCysAlaIleIlePheMetLeu 80
Db 717325 GAAATGATTCCTTCCTTCAAGTTTGGGGAATTCCTCCGGAGCTGTTATTTACTATG 717266
Qy 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValAlaGlyThr 100

```

```

Db 717265 GCTATGCGGTGTAGGACACTGCTATCCCGGAAATACGTTTTTTATGCTTCATGCGCC 717206
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 717205 GCATTCCTGTTGTTTCTTCCTGTTTGTGCTGATCATTTATCTCTGAGGAGATAGCTTG 717146
Qy 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
Db 717145 CATCTCACTCTCTGCTGATTAATTAACAAGACTCTTCTCAAGGACTTCGTGTTTTT 717086
Qy 141 ValAlaIleLeuArgAspThrPheAlaAlaPheTyrValLeuAlaGluLeuTyrGly 160
Db 717085 ATTGTATGTCCTGCTTACCTGAGATTCAGATTTATTAAGTATGTCAGAGCTGTGAGT 717026
Qy 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 717025 TCGGTTGTTCTTCGATTCGTTGTTGCTGGGACTAGCCATCAGATTAATTAATTCGAA 716966
Qy 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 716965 GCGGCGCGTTTTCAGCTCTTATCAATACAGATTAATATCTCTCAATATGCGCAGGA 716906
Qy 201 ArgAlaIleValTyrAlaSerLysLeuArg--AlaSerValSerGluGlyValAspPro 219
Db 716905 GAATATCCCTATTCAGATGGGGAACAACATTTGTTGCCACTCTTTCATGATGATTC 716846
Qy 220 Tyr--GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeu 238
Db 716845 TGGCACTGCTGAATGCTCACTGACCATGACCATGCTACT--TGTTCTGTTTAATATATG 716789
Qy 239 MetAlaSerLysTyrTrpTyrIleAsnLysAsnVal-----LeuThrAspProArgPheTyr 256
Db 716788 ATC-----TGGCTATATAGCGCGGATTCATCATTTGACTATGATTCGATC 716741
Qy 257 AspProGluLysLeuMetGlnLysGlyLysGlyAla----- 268
Db 716740 CTTCCATCTTACAGCTGCTTGGGAGAAGAGGAGGACCTACTGCTATCTTAAGGAAG 716681
Qy 269 --LysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIle 287
Db 716680 AAAAACTTAACCAAGCAAGTAAACCTTTTCTTACCTCATTCAGTCTCGTTATTTA 716621
Qy 288 LeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThr 307
Db 716620 TTAGGCTGCTATTAATGCTTATGCTTATTAATTAATGATTAATTCATTTGGAAGTCTT 716561
Qy 308 TrpLysSerGlnLeuLysLysGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327
Db 716560 TGGAGAGATCAAGTTACCGCATTTTACAGTTCACCTGATCAATTCATGATATAGT 716501
Qy 328 AsnPheSerPheTyrThrGlyValIleSerValLeuIleMetLeuPheValGlyAlaAsn 347
Db 716500 AGATACACTACCTCATTTGGCGTCTGTTTCTGATTAAGACACTGATACCTTACCGGACAG 716441
Qy 348 ValIleArgLysPheGlyTyrPheThrGlyAlaLeuValThrProValMetValLeuLeu 367
Db 716440 TGTATCCGTAAGAGGATGACCTGCGGTCTTATGTCATCTCATTTGGAATGTTAGT 716381
Qy 368 ThrGlyIleValPhePheAlaLeuValIle--PheArgAsnGlnAlaSerGlyLeuVal 386
Db 716380 TCAAGACTGCTCTTTTTCGGAACATATTTTGCAGAAAAGACATCTATATTTTGGG 716321
Qy 387 AlaMetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
Db 716320 GGAATCTTGTGAATGACACCTGCGCTGCTGAGCTGAGCTGGAGGAGGAAATGTC 716261
Qy 407 LeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGlnMetAlaTyrIlePro 426
Db 716260 CTATCCGCGGAGCAAAATTTAAGCTTTGATCAACCAAGGAATAGGCTTATATCCA 716201
Qy 427 LeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAspValAlaAlaArgPhe 446

```

Db 716200 CTTTCTCCAGAGATATAAATCATGGAAGCCGCGATTGATGCTGCTTCAAGATA 716141
Qy 447 G|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E|C|Y|S|E|I|S|E|I|G|L|Y| 466
Db 716140 GGAAGTCGAGGCTCTTTATTTATTCACAGGCGCTTGATTTCTCTCTCTGTCGA 716081
Qy 467 A|A|L|A|E|T|H|R|P|R|O|Y|L|E|U|A|I|V|A|L|L|E|U|E|U|P|H|E|I|L|E|A|L|L|E|T|R|I|P|L|E|U|A|L|S|E|R| 486
Db 716080 GCAAGTTTAAACGTATCGCCCTTCTTCATATTAATGCTGTTGATTCGGCTT 716021
Qy 487 A|A|T|R|I|L|Y|S|E|U|A|S|L|Y|S|E|U|P|H|E|U|A|L|A|G|I|N|S|E|R|A|L| 501
Db 716020 GTTGCTATATCGGTAAGAAATACCTAGAGCTGCTAGTGTAGCAACCTGAAA 715961
Qy 502 -----G|U|G|I|N|G|U|A|L|A|G|I|N|G|U| 508
Db 715960 CAACCTAAAGAACCTTCTCTTCATCGTAGCGTAGAACCCAGAACTGTAGAACAGAA 715901
Qy 509 A|S|P|S|E|R|A| 511
Db 715900 GAAATGCT 715892
RESULT 3
US-09-313-294A-5369
Sequence 5369, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Laligudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 5369
LENGTH: 280
TYPE: DNA
ORGANISM: Zea mays
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700350037H1
NAME/KEY: unsure
LOCATION: 2
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5369
Alignment Scores:
Pred. No.: 4.55e-20 Length: 280
Score: 242.00 Matches: 50
Percent Similarity: 77.33% Conservative: 8
Best Local Similarity: 66.67% Mismatches: 17
Query Match: 9.20% Indels: 0
Gaps: 0
US-09-869-433-2 (1-515) x US-09-313-294A-5369 (1-280)
Qy 402 A|A|A|I|E|G|I|N|A|E|U|L|E|U|S|E|R|Y|S|E|R|T|H|R|Y|R|A|L|E|U|P|H|E|A|S|P|S|E|R|T|H|R|Y|S|G|L|U| 421
Db 3 GCAATGCAAGACATTTTCAGTAAAGTGCAAAATACAGTCTGTTGATCCTTGCAAGAG 62
Qy 422 M|E|T|A|L|Y|R|I|E|R|P|R|O|L|E|U|A|S|P|G|I|N|G|I|N|Y|S|G|I|Y|G|A|L|A|I|L|E|A|S|P|A|L| 441
Db 63 A|T|G|C|A|T|A|C|A|T|T|C|T|T|G|A|T|G|A|G|A|T|G|A|A|G|G|T|A|A|A|G|C|G|C|T|A|T|T|G|A|T|G|T| 122
Qy 442 V|A|L|A|A|A|A|A|G|P|H|E|G|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E| 461
Db 442 V|A|L|A|A|A|A|A|G|P|H|E|G|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E| 461
Qy 123 G|T|G|T|G|C|A|A|C|C|C|T|T|G|G|G|A|A|A|T|C|G|A|G|G|T|C|T|G|A|T|C|C|G|C|A|G|C|T|T|A|T|A|T|C|T|G|C|A| 182
Qy 462 C|Y|S|G|I|S|E|R|I|E|G|I|A|L|E|U|T|H|R|P|R|O|Y|L|E|U|A|I|L|E|U| 476
Db 183 T|T|C|G|T|C|T|C|T|C|G|G|A|C|T|G|A|C|C|C|T|A|C|C|T|G|G|G|A|T|A|T|C| 227

RESULT 4
US-09-313-294A-3344
Sequence 3344, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Laligudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 3344
LENGTH: 265
TYPE: DNA
ORGANISM: Zea mays
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700611386H1
NAME/KEY: unsure
LOCATION: 28, 234
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-3344
Alignment Scores:
Pred. No.: 2.3e-11 Length: 265
Score: 168.00 Matches: 34
Percent Similarity: 76.92% Conservative: 6
Best Local Similarity: 65.38% Mismatches: 12
Query Match: 6.33% Indels: 0
Gaps: 0
US-09-869-433-2 (1-515) x US-09-313-294A-3344 (1-265)
Qy 421 G|U|N|E|R|A|T|Y|R|I|E|R|P|R|O|L|E|U|A|S|P|G|I|N|G|I|N|Y|S|G|I|Y|S|A|L|A|I|L|E|A|S|P| 440
Db 106 GAAATGCAATATATTCCTTTGATGAGATATGAGGTAAAGGAAGGACCATTTGAT 165
Qy 441 V|A|L|A|A|A|A|A|G|P|H|E|G|L|Y|S|E|R|G|Y|G|L|A|L|E|U|I|E|G|I|N|G|I|L|E|U|L|E|U|A|I|L|E| 460
Db 166 G|T|G|T|T|G|C|A|A|C|C|A|C|T|G|G|C|A|A|G|T|G|G|G|G|G|C|A|T|T|A|T|C|A|C|A|G|T|T|C|A|T|G|T|C|T|G| 225
Qy 461 I|L|E|C|Y|S|E|R|I|E|G|I|A|L|E|U|T|H|R|P|R|O|Y|L|E|U| 472
Db 226 A|C|G|T|T|C|G|N|C|C|C|T|G|C|A|A|T|T|G|A|C|T|C|A|T|A|C|T|T| 261
RESULT 5
US-09-252-991A-15706/C
Sequence 15706, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15706
LENGTH: 1365
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15706
Alignment Scores:
Pred. No.: 3.44e-06 Length: 1365

Score: 135.00 Matches: 110
 Percent Similarity: 32.25% Conservative: 58
 Best Local Similarity: 21.11% Mismatches: 163
 Query Match: 5.13% Indels: 190
 DB: 4 Gaps: 26

US-09-869-433-2 (1-515) x US-09-252-991A-15706 (1-1365)

QY 32 LeuMetPhePheCysIleThrPheAsnTyrThrValLeuArgAspThrLysAspThrLeu 51
 DB 1248 CTGGTGTCTTCGCGACACCTGAACTACATC-----GATCGCGCC 1207
 QY 52 ILeValIly-----AlaProGlySerGlyAlaGluAlaIleProPhe 65
 DB 1206 GCCCTCGCGCGTATGACCGCATCTTGCAGAAAGATGAGCTGAGCGGCAATGAGCTTAC 1147
 QY 66 -----IleLysPheThrPheValValProCysAlaIleIlePheMetLeuIleTyrAla 83
 DB 1146 GCCAACATCACTCTGTTCCAGTCCGCTACGCCATCGGCTTCATCTCCAGGCGCGC 1087
 QY 84 LysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThrProPheLeu 103
 DB 1086 TTCATGCACAAGGTCGGGCTCAAGCGCGCTTCTTCCTCGCGCTG----- 1042
 QY 104 ILePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeuHisProThr 123
 DB 1042 ----- 1042
 QY 124 GluPheAlaAspArgLeuGlnAlaIleLeuProGlyLysLeuGlyLeuValAlaIle 143
 DB 1041 -----CTGCTCTGGAGCTGGCCACCGCGGCCCGCCACCGGCTGGCGACCTCG 997
 QY 144 LeuArgAsnThrPheIleAlaIlePheTyrVal-----LeuAlaGluLeu-----TrpGly 160
 DB 996 GCGCGCGGCTCATGATCTGTCTGCTCATCTCGCGCTGACCGAGGCGCGCAATACCGC 937
 QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
 DB 936 GCTCGCGTGAAGACCCCGCTGTGTCCCGCGGGGA----- 895
 QY 181 AlaLysArgPheTyrAla-----LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAla 198
 DB 894 -----CGGGCGGTCGCGCACCGGCATCTTCAAGCGCGCACATGTGCGGCGCATGTC 841
 QY 199 SerGlyArgAlaIle-----ValThrAlaSerLysLeuArgAlaSerVal 213
 DB 840 ACCCGCGGCTGTGCTGCTGATCTGATGTCGCGGCGC----- 802
 QY 214 SerGluGlyValAspProTrpGlyIleSerLeuArgLeuMetAlaMetThrIleVal 233
 DB 801 -----TGGCAGGCGCGTTCTCTGATGTCGCGGCTG----- 769
 QY 234 SerGlyLeuValLeuMetAlaSerTyrTrpThrIleAsnLysAsnValLeuThrAspPro 253
 DB 768 -----GGCGTGTGCTGCGTGTG-----TTCGTGTGTGCGGCAAC----- 736
 QY 254 ArgPheTyrAsnProGluGluMetGlnLysGlyLysIly----- 267
 DB 735 -----TACTAACACCCCGAGAGCATCCCGGCTGAAGAGCAGCAGCTGAATATCATCCAG 679
 QY 268 -----AlaLysPLeuLysMetAsnMetLysAspSerPheLeuTyrLysAspArgSer 284
 DB 678 CAGCAGAGAGAGCCCGCGGCGCACCGCGGTGCTTCTGAGATTCGCCCGCGCGCGC 619
 QY 285 ProTyrIleLeuLeuLeuThrIleValIleAla----- 296
 DB 618 ACCTGGGCGCTTCGCGCTGCTACTCATCCGCGCGGCTTCTGTGTTCTACTCTAC 559
 QY 297 -----TyrGlyIleCysIleAsnLysIleGluValThr 307
 DB 558 TGGCTGCGCGCTTCTCAACACGACAGTACGCGCTGGCGCATCAGCGTGAACGACATG-- 502
 QY 308 TrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327

DB 502 ----- 502
 QY 328 AsnPheSerPheThrPheGlyValValSerValLeuIleMetLeu----- 342
 DB 501 -----GGCATCCCGCTGATCTCTGTTGATACCGCGCACTTCGGC 460
 QY 343 PheValGlyLysAsnValIleArgLysPheGlyTyrLeuThrGlyAlaLeuValThrPro 362
 DB 459 AGCGTGGCGCGGATCTTCG-----TCTGCTGATTCGCGCGGCGCATGCCGCG 406
 QY 363 ValMetValLeuLeuThrGlyIleValIlePhePheAlaLeu-----ValIlePhe 378
 DB 405 ATCCGCGCGGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 346
 QY 379 ArgAsnGlnAlaSerGlyLeu-----ValAlaMetPheGlyThrThrProLeuMetLeuAla 397
 DB 345 GCCGCAACCGCGCGGCTGTGATTCGCGCTG-----CTGGCGCATCGCC 301
 QY 398 ValValValGlyAlaIleGln-----AsnIleLeuSerLysSerThrLysTyr 413
 DB 300 CTGGCGGTGCGCGCGCGACGCGGTGAGCGGCAACATCTGAGC----- 256
 QY 414 AlaLeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGlnLysVal 433
 DB 255 -----CTGTGATGACTACACGCCCAAGCAGCAGTGAAGACGCTG 214
 QY 434 LysGlyLysAlaAlaIleAspValValAlaAlaArgPheGly-----LysSer 449
 DB 213 TTCGCG-----TTCGCGCGCATGTGCGCGCGCTG 184
 QY 450 GlyGlyAlaLeuIleGlnGlnGlyLeuValIleLysGlySerIleGlyAlaMetThr 469
 DB 183 GCGCGGATGTTCAAGACCG-----ATGTCGCGCGGCTGACTGACCGGACGAC 133
 QY 470 ProTyrLeuAlaValIleLeuLeuPheIleAlaIleThrLeuValSerAlaThrLys 489
 DB 132 AACACATACCGGCTGCTTCCACCATGATCCGCGCATGTTCATTCGCGCTGACCTCG 73
 QY 490 Leu 490
 DB 72 CTG 70

RESULT 6
 US-09-252-991A-15648
 ; Sequence 15648, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; ; FILE REFERENCE: 107196.136
 ; ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; ; PRIOR FILING DATE: 1999-02-18
 ; ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; ; PRIOR FILING DATE: 1998-02-18
 ; ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; ; NUMBER OF SEQ ID NOS: 33142
 ; ; SEQ ID NO 15648
 ; ; LENGTH: 1638
 ; ; TYPE: DNA
 ; ; ORGANISM: Pseudomonas aeruginosa
 ; ; US-09-252-991A-15648

Alignment Scores:
 Pred. No.: 4 6e-06 Length: 1638
 Score: 135.00 Matches: 110
 Percent Similarity: 32.25% Conservative: 58
 Best Local Similarity: 21.11% Mismatches: 163
 Query Match: 5.13% Indels: 190
 DB: 4 Gaps: 26

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 0.554 Length: 1830121
Score: 133.00 Matches: 120
Percent Similarity: 35.42% Conservative: 89
Best Local Similarity: 20.34% Mismatches: 158
Query Match: 5.06% Indels: 224
DB: 4 Gaps: 30

US-09-869-433-2 (1-515) x US-09-557-884-1 (1-1830121)

```

QY 15 PheLeuTrpProIleHisThrHisGluLeu-----LysIleValLeuPro 29
DB 1779409 TTTATTTGCGCCAGAGCAAACTGACGATTATGGAATCAAGCTAAAGTGGTATTTTGGCT 1779350
QY 30 MetPheLeuMetPhePheCysIleThrPheAsnYrThrValLeuArgAspThrIleAsp 49
DB 1779349 AATTTAGTGGTTTATGCTACTTACTTTTCAGTG----- 1779314
QY 50 ThrLeuIleValIleGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLysPheTrp 69
DB 1779313 -----TTTATAGGATTTTGA 1779299
QY 70 LeuValIleAlaProCysAlaIleIlePheMetLeuIleYrAlaIleLysLeuSerAsnIle--- 88
DB 1779298 TTTAATTT-----TTATCAGTAGAAGTCTTGGCAATATATAAA 1779263
QY 89 -----LeuSerLysGlnAlaLeuPheYr 96
DB 1779262 CTAGGCGCAAGATGAAAGAAAGAACCTGATTTACTTTCTTATCTTGGCTTGCATGTTATTT 1779203
QY 97 AlaValIleGlyThrProPhe---LeuIlePhePheAlaLeuPheProThrValIleYrPro 115
DB 1779202 GCCCGTGGATGGGGGTTGGCTGATGTTTTTGGCGTAGCAAACTTAAACCATTAAT 1779143
QY 116 LeuAlaGlyValLeuHisProThrGluProPheAlaAspArgLeuGlnAlaIleLeuProPro 135
DB 1779142 CTTTGTGACATTACAGACAGTTCTGCAGAACTAAACAAACAGAGGCTTTACTT----- 1779089
QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnYrThrPheAlaAlaPheYrValLeu 155
DB 1779088 -----CATACTGTCTTCCACTGGGAATTCACGCGGCGGACATTA--- 1779050
QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
DB 1779049 -----TATGGACACATTCCTTTAGCATTAAGCTTATTTGGGTTTCGTTATATAATTA 1778999
QY 176 Thr-----LysIle 178
DB 1778998 CCTTAGCGTGGCGCTCTGTTTATATCCTTTATTAAGATCGATTAATGAGCAAAATC 1778939
QY 179 HisGluAlaIleValArgPheYrAla-----LeuPheGlyIle----- 190
DB 1778938 GGGCGATGCAATTGATGATTATGCGCTTACTTGCACATTAATTTGGGATTTATTAACCATTA 1778879
QY 191 GlyAlaAsnIleSerLeuLeuAlaSerGly---ArgAlaIleValIleTrpAlaSerLysLeu 209
DB 1778878 GGTTTTGGTTCATACAGCTTGTGTCAGAGGCTTGAACAAATGAGTTGATTAAGTCAGAAC 1778819
QY 210 ArgAlaSerValSerGluGlyValAspProTrpGlyIleSerLeuArgLeuMetAla 229
DB 1778818 AGCTTTGGCTTACCAAGTTGGCGGTATTT-----GTTGGTGATGATGTTTACGCGGTG 1778768
QY 230 MetThrIleValSerGlyLeuValIleMetAlaSerYrYrTrpIleAsnLysAsnVal 249
DB 1778767 TTTTCTGCTATTTCTGCTGTT----- 1778747
QY 250 LeuThrAspProArgPheYrAsnAspProGluMetGlnLysGlyLysIleGlyAlaLys 269
DB 1778746 -----GGAAGGCGGTGAAA 1778732

```

```

QY 270 -----ProLysMetAsnMetLysAspSer-PheLeuYr----- 280
DB 1778731 ATATTAGTGAATTAATTAACCTTAACCTTAACCTTATTTGCTTCTTCTTTATTTCA 1778672
QY 281 ----LeuAspArgSerProYrIleLeuLeuLeuThrLeuValIleAlaYrGlyI 299
DB 1778671 GGCCCTACGTATACCTTTTATCTGCACTTATGCAATATATGGAATTT-----AT 1778621
QY 299 eCysIleAsnIleLeuIleGluValIleThrIlePheSerGlnLeuLysLeuGlnYrProAsnMe 319
DB 1778620 TTCAGC-AATTTAGTGCACACAGTTTAAAC-----TATGCTTATGA 1778577
QY 319 AsnAspYrSerGluPheMetGlyAsnPheSerPhe----- 331
DB 1778576 ACAAGAACTATCTAGCTGTTTACCGGATGACGCTGCTTATTTGGCGTGGTGCTTC 1778517
QY 332 -TrpThrGlyValIleSerValLeuIleMetLeuPheValGlyIleAsnValIleArgly 351
DB 1778516 TTGGGCTCGCTTGGGCTTATTTATTTATGCGCGTATTTCTAAAGGCGCACTATCCGTGA 1778457
QY 351 sPheGlyYrPheLeuThrGlyAlaLeuValIleProValMetValLeuLeuThrGlyIleVal 371
DB 1778456 ATTT-----ATTTTGGCGGTATTAATTTCA-----AGTTTATTTGGTATTTT 1778412
QY 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyYr 391
DB 1778411 ATGGTTT-----ACTGTTTGGTAA 1778391
QY 391 rThrProLeuMetLeu-----AlaValIleGlyAlaIleGlnAsnIleLeuSerly 409
DB 1778390 TACAGAGTATGCTAAATATGATGACATTCGCGGGCGGCGCTTGGCAATTAATTTCT-- 1778333
QY 409 sSerThrIleYrAlaLeuPheAspSerThrLysGluMetAlaYrIleProLeu----- 427
DB 1778332 -TCCCGAAGAAATTTATTAATTT-----AAATTTTAATTAATTCGCTTTAACCAAC 1778283
QY 427 ----- 427
DB 1778282 AATTAACAGGCTTTGAGCTTATTAATGATTTATTTGTTTATCACTTCAGCGGATTC 1778223
QY 428 -----AspGlnIle----- 430
DB 1778222 AGTATTATATGTTAATAATACATTCGATCTCGATTAAGTTTAAAGTTTAACTTCGCTGCGTG 1778163
QY 431 -GlnLysValIleGlyLysAlaAlaIleAspValAlaAlaArgPheGlyLysSerG 450
DB 1778162 GCAAGCCATATATGCGGAACCTTAATGTCGTTGTCGATGCTTATGATCAATCTGG 1778103
QY 450 YGly---AlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetTh 469
DB 1778102 TGGACTTGCTAATTTGCAACAAATGACATTAATTTGTC-----TT 1778061
QY 469 rProYrLeuAlaValIleLeuLeuPheIleAlaIleTrp-----LeuValSer-- 486
DB 1778060 GCTTTTGTCTTATTAATGATGTTGATGCTGTTTATGATTATGAAGGCTTAATTTGCGGA 1778001
QY 487 -----AlaThrLysLeuAsn 491
DB 1778000 TAAATAATTTTCTACTTAAAGTCAAT 1777973

```

RESULT 9

US-09-643-990A-1/C
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White

Hamilton O. Smith

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville,
 STATE: MD
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v.6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/643,990A
 FILING DATE: 23-Aug-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787
 FILING DATE: 1995-04-21
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB186P1C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-610-5790
 TELEFAX: 310-309-8439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1830121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-643-990A-1
 Alignment Scores:
 Pred. No.: 0.554 Length: 1830121
 Score: 133.00 Matches: 120
 Percent Similarity: 35.42% Conservative: 89
 Best Local Similarity: 20.34% Mismatches: 158
 Query Match: 5.06% Indels: 224
 DB: 4 Gaps: 30
 US-09-869-433-2 (1-515) x US-09-643-990A-1 (1-1830121)
 Oy 15 PheLeuTrpProIleHisThrHisGluLeu-----LysLysValLeuPro 29
 Db 1779409 TTTATTTGGCCAGAGCAAACTTCAAGCATTTAATGAATCAAGTAAAGTGTATTTTGGCT 1779350
 Oy 30 MetPheLeuMetPhePheCysIleThrPheAsnTrpThrValLeuArgAspThrLysAsp 49
 Db 1779349 AATTTTACTGGTTTATGTACTTACTTTTCAGTG----- 1779314
 Oy 50 ThrLeuIleValIGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLysPheTrp 69
 Db 1779313 -----TTTTTGGATTTTGA 1779299
 Oy 70 LeuValValProCysAlaIleIlePheMetLeuIleTyralAlaLysLeuSerAsnIle--- 88
 Db 1779298 TTTAATT-----TTATAGTAGAGTAGCTTGGCAATATATAAA 1779263
 Oy 89 -----LeuSerIlyGlnAlaLeuPheTy 96
 Db 1779262 CTAGGGCAAGATGAAGAAGACTGATTTAGTTCTTATCTTGCTTGCGCTTGCGATGTATTT 1779203
 Oy 97 AlaValGlyThrProPhe---LeuIlePhePheAlaLeuPheProThrValIleTyPro 115
 Db 1779202 GCCCGTGAATGGGGGTTGGCTGATGTGTTTTTGGCGTAGAGAAACCATTAACCATATAT 1779143
 Oy 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135

Db 1779142 CTTTCTGACATTTACAGAGGTTCTGCAGAAACATTAACAAACAGAGGCTTACTT----- 1779089
 Oy 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyValLeu 155
 Db 1779088 -----CATACTCTGTTCCACGGGAATTCACGGCTGGCAGTA--- 1779050
 Oy 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
 Db 1779049 -----TAGGCACCATTTGCTTTAGCATTTACTTATTTTGGGTTTCGTTAAATTA 1778999
 Oy 176 Thr-----LysIle 178
 Db 1778998 CTTTACCGTTGGCTCTGTTTATCTTATTAAGATCGTATTAATGCAAAATC 1778939
 Oy 179 HisGluAlaLysArgPheTyAla-----LeuPheGlyIle----- 190
 Db 1778938 GGCATGCAATTGATGTATATGCGCTTACTTGCCACATTAATTTGGATTAATACACATTA 1778879
 Oy 191 GlyAlaAsnIleSerLeuLeuAlaSerGly---ArgAlaIleValITrpAlaSerLysLeu 209
 Db 1778878 GGTTTGGTTTCATCACAGCTTGCTGCAGGCTTGCAACAATAGGTGATTAAGTCAGAAC 1778819
 Oy 210 ArgAlaSerValSerGluGlyValaAspProTrpGlyIleSerLeuArgLeuMetAla 229
 Db 1778818 AGCTTGCCCTTACAAAGTTGGCGTTAT-----CTTGCGTGAATGTTTAGCGGTG 1778768
 Oy 230 MetThrIleValSerGlyLeuValLeuMetAlaSerTyTrpTrpIleAsnLysAsnVal 249
 Db 1778767 TTTTCTGCTATTTCTGCTGTT----- 1778747
 Oy 250 LeuThrAspProArgPheTyArgAsnProGluGluMetGlnLysGlyLysGlyAlaLys 269
 Db 1778746 -----GGAAGAAGCGCTGAAA 1778732
 Oy 270 -----ProLysMetAsnMetLysAspSer-PheLeuTy----- 280
 Db 1778731 ATATTAGTGAATTAATTAACCTTAACCTTAGCATTTTGTACTCTCTGTTTAATTTCA 1778672
 Oy 281 ---LeuAspArgSerProTyIleLeuLeuLeuThrLeuLeuValIleAlaTyGlyI 299
 Db 1778671 GGCCCTTCAGTTAATCCTTTATATCGATTAGCAGCAATATTTGGTAAATT-----AT 1778621
 Oy 299 eCysIleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTrpProAsnMe 319
 Db 1778620 TTCAGC-AATTTACTGCAACTCAGTTTAAAC-----TATGCTTATGA 1778577
 Oy 319 tAsnAspTySerGluPheMetGlyAsnPheSerPhe----- 331
 Db 1778576 ACAAGAACATATCTAGCTGCTTTAGCGGATGACATGCTGCTTATTTGGCTTGCGTCTC 1778517
 Oy 332 -TrpThrGlyValaValSerValLeuIleMetLeuPheValIGlyLysAsnValIleArgGly 351
 Db 1778516 TTGGGCTCCGTTTGTGGTTTATTTATTTAGCGCGATTTCTTAAGGGGGAACATATCCGCTGA 1778457
 Oy 351 sPheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
 Db 1778456 ATTT-----AATTTTGGCGATTAGTTATTC-----AGTTTATTTGGTATTTT 1778412
 Oy 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyTh 391
 Db 1778411 ATGCTTT-----ACTGTTTGGTAA 1778391
 Oy 391 rThrProLeuMetLeu-----AlaValValaGlyAlaIleGlnAsnIleLeuSerTy 409
 Db 1778390 TACAGCAGATATGCTAATATGATGCGATGCTGCGGCGGCTGCGCAATTAATTTCT-- 1778333
 Oy 409 sSerThrLysTyAlaLeuPheAspSerThrLysGluMetAlaTyTrpLeuProLeu 427
 Db 1778332 -TCCCGAAGAAATTTTATTAATTT-----AAATTTTAATTAATTCGCTTTAACCAAC 1778283
 Oy 427 ----- 427

```

Db 1778282 AATAACAGGCTTGTGAGCTTATAGTATTTATTTTATCCTGACGCGGATTC 1778223
Qy 428 -----AspGlnGlu----- 430
Db 1778222 AGGATATTATGTTTAAATACATTCCTGCTATTAAGTTAGCTGCGCTCG 1778163
Qy 431 -GlnIysValIysGlyAlaAlaIleAspValAlaAlaArgPheGlyLysSerG1 450
Db 1778162 GCAACCCATATGTCGGGACCTTATATGCTTTCGATGTCGATGCTTATGCAATCTGG 1778103
Qy 450 YGly---AlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetTh 469
Db 1778102 TGGACTTCTTATTTGCAACATGACATTAATTTGTC-----TT 1778061
Qy 469 rProIyLeuAlaValIleLeuLeuPheIleIleAlaIleTrp-----LeuValSer-- 486
Db 1778060 GCCTTTTCCTTATGATGTTGATATGTTTATGTTATGAAAGGCTTAATGCGGA 1778001
Qy 487 -----AlaThrIysLeuAsn 491
Db 1778000 TAAAAAATATTTTCTACTAAAGTCAAT 1777973

RESULT 10
US-09-252-991A-1331
; Sequence 1331, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1331
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1331

Alignment Scores:
Pred. No.: 2.01e-05 Length: 1467
Score: 129.00 Matches: 111
Percent Similarity: 32.62% Conservative: 71
Best Local Similarity: 19.89% Mismatches: 180
Query Match: 4.90% Indels: 196
DB: 4 Gaps: 23

US-09-869-433-2 (1-515) x US-09-252-991A-1331 (1-1467)
Qy 71 ValValProCysAlaIleIle-----PheMetLeuIleTyAlaIleLysLeuSer 86
Db 7 GTCAACCCCTTCCGCTCGCGGCGGAGGTCGTGTCGTGCGGCTTCGCGCATCC 66
Qy 87 AsnIleLeuSerIys----- 91
Db 67 CGCGTCCTGCTCCGCGGCTCCCGGAGCTCAACAGCTGTGCGAGCGGCGAGTCCGGA 126
Qy 92 -----GlnAlaLeuPheTyAlaValGlyThrPro-----PheLeuIlePhe 105
Db 127 GCGGCGCGGCTCCCTGTGACGAGACCGAGATCCGATGCGCGGATTTTCTCTCTTT 186
Qy 106 PheAlaLeuPheProThrValIleTyProLeuArgPheValIleuHisProThrGluPhe 125
Db 187 CGCGCTCTCTAT-----TTC 201
Qy 126 AlaAspGlnGlnValIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArg 145
Db 202 GCTACCGCTGTGANG--CTCACCGGCTCGGCGCTGTACG----- 240

```

```

Qy 146 AsnTrpThrPheAlaAlaPheTyValIleuAlaGlu-----LeuTrpGlySer 161
Db 241 -----ACCTACCTGGGCGCTGCGCGCGGACAGATCGACGCGCTGTGGTGCGC 294
Qy 162 ValMetLeuSerLeuMetPheTrpGlyPheAla-----AsnGlu 174
Db 295 GCGGTATGAGCGGCAACTACTTGGCGCTGGTGTCGGCGGCAAGTCGGATCACGGCTG 354
Qy 175 IleThrIysIleHisGlnAlaIleAspPheTyAlaLeuPheGlyIleGlyAlaAsnIle 194
Db 355 ATCCCGCGGTCGCGCACATCCGCGCTACGTCGCTCGCGCGGGGTG----- 402
Qy 195 SerLeuLeuAlaSerGlyArgAlaIleValTrpAlaSerIysLeuArgAlaSerValSer 214
Db 403 -----GTCACCGCGCGGCTCTGGC 423
Qy 215 GlnGlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSer 234
Db 424 CATGGCTGCGCGCTCG-----CTGCGGCGCTGATCGCTTGGCGCATGATG 474
Qy 235 GlyLeuValLeuMetAlaSerTyTrp-----TrpIleAsnIysAsnValIleu 250
Db 475 GGCCTGGCGCTGATGTCACAGTACATGATGATGAGAGCTGCGCTCAAC----- 522
Qy 251 ThrAspProArgPheTyraAsnProGlnGluMetGlnIysGlyValAlaIleAspPro 270
Db 523 -----GACACAGCGCGATCCAGCCAGCGCGCGCGCTGCTTC 558
Qy 271 LysMetAsnMetIysAspSerPheLeuTyIleuAspArgSerProTyIleLeu----- 288
Db 559 GCGGCTACATNAGCGGCTTCTTACTGCTGGCTGGTGTCGGCCAGATGATCTGGTGCGG 618
Qy 289 -----LeuLeuThrLeuLeuValIleAlaTyArgIleCysIle 301
Db 619 CATCCGCACTGGGCGCGGATGCTGATGCTGATGCTGCTGCTGCTGCGCTGCTGCTG 678
Qy 302 AsnLeuIleGluValThrTrpIysSerGlnLeuIysLeuGlnIlyProAsnMetValAsp 321
Db 679 GTGCGCTGCGCTGACCCACAGATCCACCCGCGAGCGCTGCGCGCGCGCTGAA 738
Qy 322 TyrsGlnPheMetGlyAsnPhe-----SerPheTrpThrGlyValIleValSerValLeu 339
Db 739 CCAAGCTTCTTCAATCCGCGGTCGCGAGTCGCTGACACACCGATGATGCTCGGGCTG 798
Qy 340 IleMet----- 341
Db 799 GTGGTCGCTCTCTTATGCTGCGCGCTGATGCAACCAAGTTGGGATGCCCCAAC 858
Qy 342 -----LeuPheValGlyGlyAsnValIleArgIys 351
Db 859 GAGCAGGTGCGCTGATACAGGCGGCTGCATCTTCGCGGCGCTGCTGCTGCTGCTG 918
Qy 352 PheGlyTrpLeuThrGlyAla-----LeuValThrProValMetValLeu 366
Db 919 CTGGGCTGCTGTCGACCGTCGCGACCGCGCGCTGATTCGCGCGCTGCGCATCTG 978
Qy 367 LeuThrGlyIleValPhePheAlaLeuValIlePheAsnGlnAlaSerGlyLeuVal 386
Db 979 CTG-----TGCCTGTGCTGCTGCGCTG----- 1002
Qy 387 AlaMetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIle 406
Db 1003 GCATGTGTCACACAGATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062
Qy 407 LeuSerIysSerThrIysTyraAlaLeuPheAspSerThrIysGlnMetAlaTyIlePro 426
Db 1063 TTG-----CAGTTCACCTCTATCCGCTGSCAGTGGCTTTTCCAGACCAT 1110
Qy 427 LeuAspGlnGlnIysValIysGlyIysAlaAlaIleAspVal-----ValAla 443
Db 1111 GTGAGACCGACGAGCGGCTTTCGTCGACCGCATGCTGCTGTGATCCTTGCGGCTGCGC 1170

```

QY 444 AlaArgPheGlyLys---SerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuValIleCys 462
 Db 1171 GCCTGATCGGGCCGCTGCGCGCGCGCGCTGATGCGG----- 1209
 QY 463 GlySerIleGlyAlaMetThrProTyrlleuAlaValIleLeuLeuPheIleAlaIle 482
 Db 1210 ---CTGTTGGCGCTAACATGCTATATGCTTGCATCAGCGCCCTGCGCGCTGATCCTGCTC 1266
 QY 483 TrpLeuValSerAlaThrIleuAnuIleuPhe----- 494
 Db 1267 TGGCGGCTCATCCGAGAAAGTCAAGCGGCTGACCGGGTGCACGCGCGCTGCAC 1326
 QY 494 ----- 494
 Db 1327 CACGTGCGACCCCGGACAACTGACACGACTCCCGCTGTGCGCGCTGATCCGCG 1386
 QY 495 LeuAlaGlnSerAlaLeuLysGlnGlnGlnValAlaGlnGlnAspSerAlaPro 512
 Db 1387 GTAGACGAACAGCGGTCAGAGAGATGTCGATGCGGACCGGACGCGCGCG 1440

RESULT 11
 US-09-335-409-1
 / Sequence 1, Application US/09335409
 / Patent No. 6121029
 / GENERAL INFORMATION:
 / APPLICANT: Schupp, Thomas
 / APPLICANT: Ligon, James
 / APPLICANT: Molnar, Istvan
 / APPLICANT: Zirkle, Ross
 / APPLICANT: Cyr, Devon
 / APPLICANT: Geerlach, Joern
 / TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 / FILE REFERENCE: 4-30582A
 / CURRENT APPLICATION NUMBER: US/09/335,409
 / CURRENT FILING DATE: 1999-06-17
 / NUMBER OF SEQ ID NOS: 30
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 1
 / LENGTH: 68750
 / TYPE: DNA
 / ORGANISM: Sorangium cellulosum
 US-09-335-409-1

Alignment Scores:
 Pred. No.: 0.0543 Length: 68750
 Score: 122.50 Matches: 105
 Percent Similarity: 34.68% Conservative: 67
 Best Local Similarity: 21.17% Mismatches: 165
 Query Match: 4.66% Indels: 159
 DB: 3 Gaps: 21

US-09-869-433-2 (1-515) x US-09-335-409-1 (1-68750)

QY 76 IleIlePheMetLeuIleTyrlaLysLeuSerAsnIleLeuSerIleGln----- 92
 Db 3460 CTCGGGCTCATCTCGTACCGCGCGCTCCGCGAGCTCGCGCGGCGGTGGCCAG 3519
 QY 93 -----AlaLeuPheTyrlaValGlyThrProPheLeuIlePhePheAla 107
 Db 3520 CCGGAGTGTCTCGGAGAGCTCTTCGCGCGGCTC-----GTG 3555
 QY 108 LeuPheProThrValIleTyrlProLeuArgAspValLeuHisProThrGlnPheAlaAsp 127
 Db 3556 CTGGGCGCTCGTCCGCGCGCGCTCGCGCGCGGATTCATCGAGCCCTTTCAGAGAG 3615
 QY 128 ArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrp 147
 Db 3616 CCGGCGGTGCGGTGCTGCTCGGCGCATC----- 3645
 QY 148 ThrPheAlaAlaPheTyrlValLeuAlaGlnLeuTrpGlySerValMetLeuSerLeuMet 167
 Db 3646 -----TCCTGATAGCGCGCTCTCTCTGCTGCTG 3675

QY 168 PheTrpGlyPheAlaAsnGlnIleThrLysIleHisGlnAlaLysArgPheTyrlaLeu 187
 Db 3676 ATGGGCGCATCGAGCTGAGCTGAGCATCTCTGCGCAAGAGGCGCGCGCGCTC 3735
 QY 188 PheGlyIleGlyAlaAsnIleSerIleuLeuAlaSerGlyArgAlaIleValTrpAlaSer 207
 Db 3736 TCGGCGCTCGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3780
 QY 208 LysLeuAlaAlaSerValSerGlnGlyValAspProTrpGlyIleSerLeuArgLeuLeu 227
 Db 3781 ---TTCGCGGCTCGTGTCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3837
 QY 228 MetAlaMetThrIleValSerGlyLeuVal-----LeuMet 239
 Db 3838 CTCTCGGAGAGCGCGCTGATCGGATCGGAGAGCTGATCGAGCGGAGTGCATGCGC 3897
 QY 240 AlaSerTyrlTrpTrpIleAsn-----LysAsnValLeuThrAspProArgPheTyrlAsn 257
 Db 3898 CGCAGCTATGCGCAGGTGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3957
 QY 258 ProGlnGlnMetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSer 277
 Db 3958 CTCTGCGGATGACGTCGTGACGTCAGCTACGCGCGCGCGCGCGCGCGCGCGCGCG 4017
 QY 278 PheLeuTyrlLeuAspArgSerProTyrlleuLeuLeuThrLeuValIleAlaTrp 297
 Db 4018 GCGCTCCG-----GCGAGCGGATCTTGTCTGTCATAGTGTCTGTC----- 4059
 QY 298 GlyIleCysIleAsnLeuIleGlnValThrTrp-----Lys 309
 Db 4060 GGGCGCGCGCTACCCACCTCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4119
 QY 310 SerGlnLeuLysLeuGlnTyrlProAsnMetAsnAspTyrlSerGlnPheMetGlyAsnPhe 329
 Db 4120 GACAGAGGTGCGCTC----- 4134
 QY 330 SerPheTrpThrGlyValIleValSerValIleMetLeuPheValGlyLysAsnValIle 349
 Db 4135 -----GTCTGCTCTCACGTTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 4170
 QY 350 ArgLysPheGlyTyrl-----LeuThrGlyAlaLeuValThrProValMetVal----- 365
 Db 4171 CAGCGGCTCGCGCTGACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4230
 QY 366 -----LeuLeuThrGlyIle----- 370
 Db 4231 GCTCTCGCACCAACCGCGCTCTCTCTCGACGCGCGCGCGCGCGCGCGCGCGCGCG 4290
 QY 371 -----ValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----Val 386
 Db 4291 GCGCTGTGTCTCTCTCTCTCGCGCGCATGCGCGCTGACGCTGCGACGCTGCGACGCG 4350
 QY 387 AlaMetPheGlyThrThrProLeuMetLeuAla-----ValValVal 400
 Db 4351 GCGGCGTGGGAGAGGTGCGCTTCTGTCGCGACCGGAGCGGCGGAGAGTGTGCCCC 4410
 QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrTyrlaLeuPheAspSerThrLys 420
 Db 4411 GCGGCGCTCGCGCGCGCGCTCGC-----GCGCTCAGGCGGAGCGAGCGG 4455
 QY 421 GluMetAlaTyrlleProLeuAspGlnGlnLysValLysGlyValAlaAlaIleAsp 440
 Db 4456 GCGCTCGTGGGCGGTGGCGCTTAACATG-----AAGGCGGACGAGAC 4497
 QY 441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuVal 460
 Db 4498 CTGATCGTGGCGAGTGTGCG-----GTGAGCTGCGGCTCTCTC--- 4536
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrlleuAlaValIleLeuLeuPheIle 480
 Db 4537 -----TCCACGAGGCTTATACGATGACCGCGTGTGCGGTGTGACGAGTGCAC 4587
 QY 481 Ala-----IleTrpLeuValSerAlaThrLysLeuAnuLysLeuPheLeu 495


```

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Alignment Scores:
Pred. No.: 0.0543 Length: 68750
Score: 122.50 Matches: 105
Percent Similarity: 34.68% Conservative: 67
Best Local Similarity: 21.17% Mismatches: 165
Query Match: 4.66% Indels: 159
DB: Gaps: 21

US-09-869-433-2 (1-515) x US-09-567-969-1 (1-68750)
Qy 76 I l e l l e p h e t l e u l e u t y r a l a l y l e u s e r a n l l e u s e r l y s g l n ----- 92
Db 3460 C T G C G C T C A T C C T G A C G C G C G C C T C C G C C A G C T C G G C G C G C G C C A G 3519
Qy 93 ----- A l a l e u p h e t y r a l a v a l g l y t h r p h e u l e u l e p h e a l a 107
Db 3520 C C G A G G T G C T C G G G A G C T T C C G G G C G T C ----- G T G 3555
Qy 108 L e u p h e p r o t h r v a l l e t y r p r o l e u a r g a p v a l l e u h i s p r o t h r g l u p h e a l a s p 127
Db 3556 C T G G C C C T C C C G C T C G C G C G C G C G C C G G G T T C C A G C C C T T C C A G G A G 3615
Qy 128 A r g l e u d i n l a l e u l e u p r o p r o g l y l e u l e u d i l y l e u v a l a l l e l e u a r g a n t r p 147
Db 3616 C C G G C G T C G G G T C G T G C T C T C G G G A T C ----- 3645
Qy 148 T h r p h e a l a a l a p h e t y r v a l l e u a l a g l u l e u t r p g l y s e r v a l m e t l e u s e r l e u m e t 167
Db 3646 ----- T C C T G A T A G G C G C G C T C T C T C G T C G T 3675
Qy 168 P h e t r p l y p h e a l a a n g l u l e t h r l y s l l e h i s g l u a l a l y a r g p h e t y r a l a l e u 187
Db 3676 A T G C C G C G T C G A C G T G C A C T G G C A T C C T G C C A A G A G A G C G C C C C G G G C G C T C 3735
Qy 188 P h e g l y l l e g l y a l a a n l l e s e r l e u l e u a l a s e r l y a r g a l l e v a l t r p a l a s e r 207
Db 3736 T C G G C G T C G A G C G A T C G C G C C C C C G C T C G G G A G C G C G C C ----- 3780
Qy 208 L y s l e u a r g a l a s e r v a l s e r g l u g l y v a l a s p p r o t r p l y l e s e r l e u a r g l e u 227
Db 3781 ----- T T C T C G C G C T C G T C G A T C G C C C C T T C C A G C G C C T T C T C G G A T G T G 3837
Qy 228 M e t a l a m e t h r l e v a l s e r g l y l e u v a l ----- L e u m e t 239
Db 3838 C T C T C G G T G A C G C G C T C A G G T A T G C G A A G G T G C T G A T C A G C G C A G T C A T G C G C 3897
Qy 240 A l a s e r t y r t r p l l e a n ----- L y s a n v a l l e u t h r a p p r o a r g p h e t y r a s n 257
Db 3898 C G C A G C T A T G C A G G T G A C C T C G C G C G G G T G T G A C G A G G T G C T G C T G G A T G 3957
Qy 258 P r o g l u m e t g l n y s g l y l y l a l a l a s p r o l y s m e t a m m e t l y a a s p s e r 277

```

```

Db 3958 C T G C G C G A T G A C G T C T G A C T A C G C G C G C C C G C T G G C G G T C C C G A G C 4017
Qy 278 P h e l e u t y r l e u a s p a r g s e r p r o t r p l l e u l e u t h r l e u e u v a l l e a l a t y r 297
Db 4018 G C G C T C T C G ----- G G A G C G A T T C T T G C T G T C A T G A T G T C T C T C 4059
Qy 298 G l y l l e c y s l l e a n l e u l l e g l u a l t h r t r p ----- L y s 309
Db 4060 G G C G C G C G C T C A C C A C C T C G G A T G C G T G G T G G C C A C C G A C C G C G C T C C A G 4119
Qy 310 S e r g l u l e u l s l e u g l n t y r p r o a n m e t a s n a s p t y s e r g l u p h e t g l y a s n p h e 329
Db 4120 G A C A G G T G C G C T C ----- 4134
Qy 330 S e r p h e t r p h r a l y a l a s e r v a l l e u l l e m e t l e u p h e a l a g l y a s n v a l l e 349
Db 4135 ----- G T C C T G T C T C A C G T T C T C G C G C G C G C G C T G A C G 4170
Qy 350 A r g l y p h e g l y t r p ----- L e u t h n g l y a l a l e u v a l t h r p r o v a l m e t v a l ----- 365
Db 4171 C A C G G C T C G C C T G C A C C G C C T G C T C G C G C G C G C G C T T G C G C G T G C T C A A G C 4230
Qy 366 ----- L e u e u t h n g l y l e ----- 370
Db 4231 G T C T C T C G C A C C A C C G C C C T C T C T C A C G C G C G T G A G A C G C T G G C G G C C T C T T C 4290
Qy 371 ----- V a l p h e p h e a l a l e u v a l l e p h e a r a s n g l a l a s e r g l y l e u ----- V a l 386
Db 4291 G C G C T G T G T T C T C T C T C G G G C A T G C G G T G C A C G T T C G A C G T G C G A C G C G C G C 4350
Qy 387 A l a m e t p h e g l y t h r t h r p r o l e u m e t l e u a l a ----- V a l v a l a l 400
Db 4351 G C G C G T G G G G A C G T C G C G T T C T G C T G C C A C C G C G C G C G A A G T C T C C C 4410
Qy 401 G l y a l l e g l a n l l e u s e r l y s s e r t h r l y s t y r a l a l e u p h e a s e r t h r l y s 420
Db 4411 G C G C G C T C G C G C G C G C G C T C G C ----- G G G C T C A G G C G A G C A G C G C 4455
Qy 421 G l u m e t a l a r t l l e p r o l e u a s p g l n g l u l y s v a l l y s a l a l a l l e a s p 440
Db 4456 G C C C T C T G C G G T G G C C T A C A T G ----- A A G G C G C G A C C G A C 4497
Qy 441 V a l v a l a l a l a r g p h e g l y l y s s e r g l y l a l e u l l e g l n g l n g l y l e u v a l 460
Db 4498 C T A T C T C G C G A T C T C G C ----- G T C A A C T C G G C T C T C T C 4536
Qy 461 I l e c y s g l y s e r l l e g l a m e t t h r p r o t r p l e u a l a v a l l e l e u e u p h e l l e 480
Db 4537 ----- T C C A A C G A G C T T A D A G A T A G C C G T C G C G C T C G C G T C A C G G T A C C 4587
Qy 481 A l a ----- I l e t r p l e u a l s e r a l t h r l y s l e u s n l y l e u p h e l e u 495
Db 4588 G C C T C A C C C G C G C T C T C A T C G T C G C A G ----- 4617
Qy 496 A l a g l n s e r a l a l e u l y s g l u g l n g l u a l a l a g l n g l u s p e r a l a 511
Db 4618 ----- A A A A G G C G C C T C G A C G A G A G A G T C G C T 4650

RESULT 14
US-09-568-480-1
Sequence 1, Application US/09568480
Patent No. 6355458
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,480
CURRENT FILING DATE: 2000-05-10

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 25, 2003, 10:17:18 ; Search time 5701 Seconds
(without alignments)
3695.572 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKTEBKRFGLRSLFPIH.....AQSLKEQVNAQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2886711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+ .p2n.model -DEV-xlh
-O=/cgn2.1/USPTO/US09869433/runat_24112003_171459_18120/app.query.fasta.1.711
-DB=genemb1 -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCH=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09869433 @CGN 1.1 3635 @runat_24112003_171459_18120 -NCPU=6 -ICPU=3
-NO MAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.*
1: gp_da.*
2: gp_hcg.*
3: gp_in.*
4: gp_om.*
5: gp_ov.*
6: gp_pat.*
7: gp_ph.*
8: gp_pl.*
9: gp_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_in.*
18: em_in.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_ats.*
28: em_un.*

29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rdd.*
36: em_htg_mam.*
37: em_htg_vtc.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2617	99.5	11421	1 AE001619	AE001619 Chlamydia
2	2617	99.5	13389	1 AE002202	AE002202 Chlamydia
3	2617	99.5	110000	6 AR310754_03	Continuation (4 of
4	2617	99.5	110000	6 AR310754_04	Continuation (5 of
5	2610	99.2	300650	1 AP002546	AP002546 Chlamydia
6	2242.5	85.3	300512	1 AE016995	AE016995 Chlamydia
7	2147	81.6	14168	1 AE001281	AE001281 Chlamydia
8	2144	81.5	1587	1 TAJ10586	TAJ10586 Chlamydia
9	2143	81.5	23773	1 AE002301	AE002301 Chlamydia
10	1379.5	52.5	1850	6 AX654307	AX654307 Sequence
11	1379.5	52.5	1851	6 AX660386	AX660386 Sequence
12	1338.5	50.9	1872	6 AX654737	AX654737 Sequence
13	1333	50.7	1875	6 AX505999	AX505999 Sequence
14	1333	50.7	1875	6 AX651697	AX651697 Sequence
15	1333	50.7	2110	8 AF428316	AF428316 Arabidops
16	1331	50.6	1933	6 AX660088	AX660088 Sequence
17	1330.5	50.6	1896	8 AY128844	AY128844 Arabidops
18	1330.5	50.6	2090	8 AY081350	AY081350 Arabidops
19	1330.5	50.6	2146	8 AY084374	AY084374 Arabidops
20	1330	50.6	2181	8 AY045903	AY045903 Arabidops
21	1329	50.5	2151	8 AY045903	AY045903 Arabidops
22	1306	49.7	1896	6 AX098423	AX098423 Sequence
23	1306	49.7	2300	8 STPLATDIR	STPLATDIR
24	1302.5	49.5	2126	8 GSU251356	GSU251356 Citrus hy
25	1302	49.5	2122	8 AY098893	AY098893 Citrus hy
26	1260	47.9	1770	6 AX098422	AX098422 Sequence
27	1198	45.6	91001	8 AC013453	AC013453 Arabidops
28	1187	45.1	132990	8 AC018848	AC018848 Arabidops
29	1157.5	44.0	2139	8 AY045903	AY045903 Arabidops
30	1107	42.1	174617	2 AP003977	AP003977 Oryza sat
31	1088	41.4	1503	1 RMO507299	RMO507299 Oryza sat
32	1079.5	41.0	1500	1 CC441310	CC441310 Caedibact
33	1079	41.0	1497	1 RRI507300	RRI507300 Rickettsi
34	1079	41.0	148018	2 AE003974	AE003974 Oryza sat
35	1076	40.9	14117	1 AE008575	AE008575 Rickettsi
36	1073	40.8	1497	1 RRY507301	RRY507301 Rickettsi
37	1056	40.2	1505	12 AF087957	AF087957 Synchetic
38	1056	40.2	1604	1 RIFATPDTA	RIFATPDTA
39	1056	39.0	282610	1 RPYX01	RPYX01
40	1025.5	39.0	1823	6 AX100499	AX100499 Sequence
41	1025.5	39.0	10163	1 AE001646	AE001646 Chlamydia
42	1025.5	39.0	10407	1 AE002174	AE002174 Chlamydia
43	1025.5	39.0	110000	6 AR310754_07	Continuation (8 of
44	1025.5	39.0	300550	1 AP002547	AP002547 Chlamydia
45	1020	38.8	1900	1 AY120885	AY120885 Holospora

RESULT 1

ALIGNMENTS

AE001619 11421 bp DNA linear BCT 01-DEC-2000
LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.
ACCESSION AE001619 AE001363
VERSION AE001619.1 GI:4376620
KEYWORDS
SOURCE Chlamydia pneumoniae CML029
ORGANISM Chlamydia pneumoniae CML029
REFERENCE Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS 1 (bases 1 to 11421)
Kalan, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,
Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PubMed 10192388
REFERENCE 2 (bases 1 to 11421)
Kalan, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Direct Submision
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Baril Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. 11421
/organism="Chlamydia pneumoniae CML029"
/mol_type="genomic DNA"
/strain="CML029"
/db_xref="taxon:115713"
/gene="yCgd"
/note="synonym: Cpn0346"
complement (95. .1069)
/gene="yCgd"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAD18490.1"
/db_xref="GI:4376621"
/translation="MALGPSYGVGSFPOFSPFSPRLFGSLFTGSLYINDIQIVT
LAISGAFAGFLVLRKAMANNVSHVTLGVLCVCIPTHQTLTSLGTLTAMMA
TAMLTGFLVTRNFVKSSESTALVFLSLVLLVEMTKNAHIGTELVLGNAD
SLTKEDIPVTVILANAVITIFARSLVCSFDSVPASLSGIPRLVDYLIIFQLSA
CLVGAFKAGVLMALAFILIPSLIAKVIKSIKSIAMSLVPSIGTAFAPASSRAIL
SAVDGLSTSGSVFLTMWYIVKFIISVFRGYPSKNEPKISEKSSQ"
complement (1069. .2415)
/gene="yCgc"
/note="synonym: Cpn0347"
complement (1069. .2415)
/gene="yCgc"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAD18491.1"
/db_xref="GI:4376622"
/translation="MLSCVFSPTIFLSSFLAVTLICMTTALMGTLISKOPLESSE
SHASVGLVGLMAOVFSLQASIFVILFEGCASVFGYHIVLGVCKLHMSAL
CPVLYPFAIGLILASVYKESPTLYNRINATLQCAATIGETLTAIYPCASLPA
LMMYRQIVVTFEDKDFATCGKTLVLEALSILISLVISGVASGIVLISAFVA
PSLGRQSLDRSLTLLISAFPGISGLAGSYISVAFICRAIIGQAAVPLVLPGLV
VICAGLAVGLCLFSPKSGWVIFVRKIKFSSKDEHLKVFHSHNRLENTSVRD
FVCSYKQOYELCPKPPRWVOILEMRGVKKEODYRLTKKGRSEALRLVAHLMW
SYLVNSLPSKSVHLEAIEIETHVLTBELDHTLTILNDPCVDPHROIIPKKKEV"
complement (2418. .3197)
/gene="yCg8_1"
/note="synonym: Cpn0348"
complement (2418. .3197)
/gene="yCg8_1"
/codon_start=1
/transl_table=11
/product="rRNA methylase"
/protein_id="AAD18492.1"
/db_xref="GI:4376623"
/translation="MNVDKDFWVSVHNLGVYEHAAVLVHISFLGKSLTALLGPNG

AGKSLTKASLGLIKPSSGTVYFFNOKKRYORLAVPORASVDMPFMTVLALAM
GCYAKGMMWRISDSDRRAFHILERVLESDVADROGLSGGQOQAFILBALMOKA
DLYMDLEFSAIDMASFKTSYGVLOELBDOKTIVYVHHDLSHRQLDPHVVLNKLRL
ICCGPTDECLNGDITPQYGCIELEBQTLKSKQKQSC"
complement (3181. .4164)
/gene="yCgA"
/note="synonym: Cpn0349"
complement (3181. .4164)
/gene="yCgA"
/codon_start=1
/transl_table=11
/product="Solute Protein Binding Family"
/protein_id="AAD18493.1"
/db_xref="GI:4376624"
/translation="WDAMGYIFKYMRIFCFVACGTRPGCTSGFONANSRPLISM
NRMTHDCYERVGNRLATAVILKSLDPHATEMVKGDOKTAGSAVIFCNGLGHTL
SLRKHLENNPNSVKLERLARGAFLPDEEGICDPHIMPLUSTWKEVAITEVLEIL
KEPWSAEPKNSBELVCEMSILDSMAQCSITPENLRVLSGHNAFSEYTRLAT
PEEYASGAMRSRCSPEGLSPEAOISVDIAVAVDINEHDVSVFPEDTNOALAKK
IVSILKKSILVRLAQKPLSDNDVDFSTFKHVCITTELGVALEQCR"
4575. .5081
/gene="Cpn0350"
4575. .5081
/gene="Cpn0350"
/codon_start=1
/transl_table=11
/product="CT066 hypothetical protein"
/protein_id="AAD18494.1"
/db_xref="GI:4376625"
/translation="MAVEQSHIKEIEIKLIGKAIKRYCNGKENDLCRYLPGPSGYMH
HFTLKKMSAPBQLMKLTKTFLIESEPTPIPKPAPRSGSKRRDIPNFKTIDIER
VLELROVGDQDLARFSPKKPLSLKRELIRSRNGIVSVELMNAVYEAQVASSPN
LEVTSPTV"
5282. .6829
/gene="ad_1"
/note="synonym: Cpn0351"
5282. .6829
/gene="ad_1"
/codon_start=1
/transl_table=11
/product="ADP/ATP Translocase"
/protein_id="AAD18495.1"
/db_xref="GI:4376626"
/translation="MTKTEKRFGLKSLFPLHITHLKKVLPWLEFCTFNVTVL
PDTKPTLVAGRSGGAAIPRIKMWVPCAIIIMLYAKLSNLISQALFYAATGP
LIFRDPPTVYIPRDVLPHPEDRQALIPBULSGVALIRWTRAPVLAELMG
SVMLSLMFWGFANEITKIHAKRYALFGIGANTSLASGATVWASKLRASVSEYD
PMGISLRLLMAMTIVSGVLNASYWLNKVLTDPFRYNPEKQKSGKGAAPKXNMED
SFLYLARSPLYLLALVIAVGCINLIEVYTKSQLXLOYPNNMDYSEFMGNFSGMTG
VVSVLMLFVGANVIRKGMVLTGALVTPVAVLTGIVFALVIRNOASGLVAMEGTT
PLMLAVVGAIONILSKSTKVALPDSIKEMAYIPIDDOEKYKGAALDVAARSGSG
GALIQGLVITCGSIGMTPLVAVILFIIAIVLSATKLKPLAOSALKEQVAOE
DSAPASS"
6968. .8227
/gene="Cpn0352"
6968. .8227
/gene="Cpn0352"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18496.1"
/db_xref="GI:4376627"
/translation="MKSPKLTIVPHIFLKNDCBGRSCSKIRTIABLIIGLVALVS
ALSFVLAAPSTAIIGTLAALVILITLVALLKSKVLPINELOKTIYRYRK
EVFVFTVTHSLVTVNELKIFINCWKSGTDLDPNLRKAEAGIDILKSIDLTPPEEB
ILLONCPLYWLSHFIDKTESVAGEIGLNKQKVGGLGPIAFHKGVTITPISRYRPL
TLISESQKFLYSKASKNOWDSVKTCTEIKFELPHNMIFRDVOGIGSQFLPFS
HGITWBOAKIOLINPDNMKMLCOFDAQGCSWATRGFLNTEETMNPDPVSNVEPT
VNRTWKEKLVLEKYESPMHPSASLVOKICVYTHHQLKRWQVVRNITSQWTS
LPQYAFHAGQYKLEKKTLESSLPKSSL"
8287. .8532
/gene="Cpn0353"
8287. .8532

```

/gene="CPn0353"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8497.1"
/db_xref="GI:4376628"
/translation="MRRPNKSLIFIKIRISQSGCKEOKARPFKKSITSSIVLL
LEAFENPSSIIQNNFNKPNKKNKINIRIFVKFTI"
8561..9904
/gene="CPn0354"
8561..9904
/gene="CPn0354"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8498.1"
/db_xref="GI:4376629"
/translation="MSNITSPVIGNNSCNVYPKLNSTTIHIVISAILLGALIAFL
CVAPVYIISGALIGLILALIGLIGLIIKIPMISKEQVPEQLNIRIRAHYK
FVDFVSEAKPNLKDLSFDILNOLSENGSTNVVSELOOKIDTFGIRAKNE
VRTASIKRLSSAASRPLPBLPKILQKVPFPGFISNGSVVILHVKIKGSL
BEDLSDIYKPEMLPTTWLIDPFRPTNSIUNHLIVLARLIRDVQHLKVALNGE
WNLNHSULNTWKQQLFAKYHAAVSYHLSQPSQEDFYNLILCIFKHYSWKQMSL
IKTVPADLWMLCCLTDHTGRPDMEFASLIGTLVYQGLIHKSEBAFLSLTSLSD
QFTIRROSTNIAMFLNLTATNHSFRLPPIVHPLKRSVFSQPEDESSLIG"
complement(9982..11283)
/gene="CPn0355"
complement(9982..11283)
/gene="CPn0355"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8499.1"
/db_xref="GI:4376630"
/translation="MSMTIVPHALFKNHCCHSTFPLSRITVAIASLFCIGALAA
LGCIAPVSYIVGSVLAIFVILSLIVLILFEKCLPTPRIIPDRFHVDEAG
LSIAFVREOOVTLAEFROFSTALCNISPEKIKQPSLELSRVESFGISRLAGLE
KNMNPIDEDLSOTCPILVYLQKISAGDPQVCRDLYPRCCYGYMGLPYSTAKT
IFCKETHHIIQQITKEDVYLKAKLDEKMDTEYKAIIVERITYTTTARGLITRAGS
LITETIKELLSLHGYSFDQLITQLPDPDAMDWCIVDNSTAYNLQCALVALS
SOMLDESSIDFDNLGLVYIDLKEAVQAFASDEKELGKFLRHLSSVSRKLES
VLRQGLRIRLLENHARARVYDVFVGVAFHRSKTSIFPD"
BASE COUNT 3791 a 2376 c 1936 g 3318 t
ORIGIN
Alignment Scores:
Pred. No.: 5.5e-222 Length: 11421
Score: 2617.00 Matches: 513
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 2
Query Match: 99.51% Indels: 0
DB: 1 Gaps: 0
US-09-869-433-2 (1-515) x AE001619 (1-11421)
QY 1 MetThrIysThrGluGluYrsProPhcGlyLysIleuAgsSerPheLeuTrpProIleHis 20
DB 5282 ATGACAAAACCGAAGAAAACCTTTGGAAAATGGCTCTTTCTTGCGCCGATCAT 5341
QY 21 ThrHisGluLeuYsLysValIleuProMetPheLeuMetPhePheCysIleThrPheAsn 40
DB 5342 ACTCACAGAGCTAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGATTAAC 5401
QY 41 TyrThrValIleuAgsPthrLysAspThrIleuIleValGlyAlaProGlySerGlyAla 60
DB 5402 TATACGATGTATCCGATACAAAGACACTCTTAATGGAGAGCTCCGTGTTGCGTCA 5461
QY 61 GluAlaIleProPheIleLysPheTrpLeuValIleProCysAlaIleIlePheMetLeu 80
DB 5462 GAGGCAATACCTTTCATCACTTTTGCTGTGTGCTCCCTGTGCTATATCTTATGCTT 5521
QY 81 IleTyrAlaLysIleuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100

```

```

DB 5522 ATTATGCAAGCTAAGTAATATTTTAAGTAAGACGGCTTATTTATGACGTGGAAACG 5581
QY 101 ProPheIleuIlePhePheAlaLeuPheProThrValIleTyrProLeuAgsPheValLeu 120
DB 5582 CCCCTTTTAATTTTCTTGGCCGTTGCCACGATGTAATTAATCCGACACCGAATGTTTAA 5641
QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
DB 5642 CATCTACAGAAATTTGCTGACCGTTTACAGGCCATCTTACCTCAGGATGCTAGGATC 5701
QY 141 ValAlaIleLeuAgsPthrIlePheAlaIlePheTyrValIleuAgsIleuLeuTrpGly 160
DB 5702 GTTGCCATCTTAAGAAACGTGACATTTGCTGCAATTTATGTACTGCTGCAACTAGGGGA 5761
QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
DB 5762 AGGTCATGTATCTCTAATGTTCTGGGATTTGCTAATTAATTCAAAATTCACAGAA 5821
QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyValaAsnIleSerLeuLeuAlaSerGly 200
DB 5822 GCAAAGCTTTCTACGCTCTTTTCGGTATGAGCTATATTTCTTACTAGCTTCGTGT 5881
QY 201 ArgAlaIleValIlePheAlaSerLysLeuArgAlaSerValSerGluGlyValaAspProTrp 220
DB 5882 CGTGCAATGTTTGGGCTTCAAGGTGAGAGCTTCGTTCTGAAGGTGATAGATCCTTGG 5941
QY 221 GlyIleSerLeuAgsLeuLeuMetAlaMetThrIleValSerGlyLeuValIleuMetAla 240
DB 5942 GGAATTCCTTACGCTCTTTTGATGGCTATGACATGTAATGTAATGTAATGTAATG 6001
QY 241 SerTyrTrpTrpIleAsnLysAsnValIleuThrAspProArgPheTyrAsnProGluGlu 260
DB 6002 AGTTACTGATGATCAATTAAGACGATTAAGACGATTCCTGCTTCTTAATCAAGAA 6061
QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
DB 6062 ATCAAAAGGGAAGAAAAGGTGCTAAACCTAAATGAATGAAGATACCTCTCTAT 6121
QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
DB 6122 CTTCGTAAGATCTCTTAATATCTTTATATGCTCTCTGCTTATGCTTATGATATTGC 6181
QY 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnIleProAsnMetAsn 320
DB 6182 ATTAACCTTAATCAAGAGACTTGAAGAAAGTCACTGAACTGCAATTAATTAATGAAT 6241
QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValIleSerValIleuIle 340
DB 6242 GACTATATGATGATTAATGAGGAATCTCTCTTGGACTGGCGTATGATCGTATCTTATC 6301
QY 341 MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTyrPheThrGlyAlaLeuVal 360
DB 6302 ATGCTATTGTTGTGTGTAACGTCAATCTGAATTTGATGATGTTAACTGAGGCCCTGATC 6361
QY 361 ThrProValMetValIleuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
DB 6362 ACTCTGTCATGATGTTCTCTTAACAGGATGCTTTCTGCTCTGTTATCTTTAGAAC 6421
QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValIleVal 400
DB 6422 CAAAGCTTCGGGTGGTGGCTATGTTCCGTACCACTCTCTCATGCTAGCTGCGTTGTC 6481
QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
DB 6482 GAGACTATACAGAAATATCTTTGGAATCCACAAAATACCTCTCTTTCGACTCAACTAAA 6541
QY 421 GluMetAlaTyrIleProLeuAspGlnGluGlnLysValIleGlyLysAlaIleAsp 440
DB 6542 GAAATGGCTTAATCCCTCTTGACCAAGACAAAAGTCAAAAGGTGAAGGTGCTATATGAT 6601
QY 441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuVal 460
DB 6602 GTAGTGCTGCTCCGCTTCGGAATAATCAGAGAGGCTTTAATCAACAAGTTTGCTGTT 6661

```

QY 461 IIleCyGgISerIleGlyAlaMeChPrProTyrLeuAlaValIleLeuPheIle 480
Db 6662 ATCTGTGAAGATTGGAGCTATGACCCCTTATCTTGAGATTCCTTCTTCTACT 6721
QY 481 AAlaIleTLeuValSerAlaThrLyLeuAnuLyLeuPheLeuAlaGlnSerAlaLeu 500
Db 6722 GCTATTGCTGTTGTTCTTCGCAACTAGTTAAACAACCTATTCTTAGCGCACTGCTCTT 6781
QY 501 LySGluGluGluValAlaGlnGluGluSerAlaProAlaSerSer 515
Db 6782 AAGAACAAGAGCTGCTCAAGAAGATTCACTCTGCTTCTTCA 6826
RESULT 2
LOCUS AE002202 13389 bp DNA linear ECT 30-MAY-2000
DEFINITION Chlamydomophila pneumoniae AR39, section 33 of 94 of the complete genome.
ACCESSION AE002202 AE002161
VERSION AE002202.2 GI:8163425
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae AR39
REFERENCE Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.
AUTHORS 1 (bases 1 to 13389)
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uterback, T., Berry, K., Bass, S., Linner, K., Weidman, J., Khouiri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
TITLE Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
JOURNALS Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 13389)
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uterback, T., Berry, K., Bass, S., Linner, K., Weidman, J., Khouiri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189324.
FEATURES
SOURCE
Location/Qualifiers
1..13389
/organism="Chlamydomophila pneumoniae AR39"
/mol_type="Genomic DNA"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
163..339
/gene="CP0400"
163..339
/gene="CP0400"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038245.1"
/db_xref="GI:7189325"
/translation="MLSVSKMLPTQDSIERETDSKDKKIFTTYTSSKYLACHPFS HLDGHRKHISIGV"
597..1448
/gene="CP0401"
597..1448
/gene="CP0401"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

gene
CDS
/product="hypothetical protein"
/protein_id="AA038246.1"
/db_xref="GI:7189326"
/translation="MNIQPEYRNTOVNSQATQFSCQALSLITYSVAAVATATVAVCSQSLSTLEIGTALVLSLIFASAMFNIYKRODEPKLLIPKIMELIQEYPSI LVPSIROEVSIVEIHILLISLKNKTNFADAPYVLDKQFGEIKEDYHPSLDPNF EEILIQCPHMGRLVPMVSDVTGTYGYWCGPGLYENAPSLFERSLSLLKKI SPSPFLIEDGLKNTWSSSELVQIRNQLPTRYYADKEVDEALNADYEQFDSLLHL IFSHKLS"
1455..1808
/gene="CP0402"
1455..1808
/gene="CP0402"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038247.1"
/db_xref="GI:7189327"
/translation="MOLFOYNESGMDWLCDPDSQGBQSLVGLIHSSMALYEAK EQYLPVRSLLTWBELIMQULSKPTKGVAKDLCNVFEHGFQRFQYGLSDLNQRF ENTFLNPKYHLDR"
1842..1958
/gene="CP0403"
1842..1958
/gene="CP0403"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038248.1"
/db_xref="GI:7189328"
/translation="MKSLSVRSAPFSGIYKSLPNLKKELPFMGFFITNS"
2019..3326
/gene="CP0404"
2019..3326
/gene="CP0404"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038249.1"
/db_xref="GI:7189329"
/translation="MAMSWTIVPHALFKHCECHSPPLSSRTIYRIANSLFCIGAL AALGCLAPVSYVGSVLAIFAVILSLVILALIFBKLLPPTPRIIPDFTVIDEA YGSLSAFVSEQVTLAEFRQSTALCNLSPEKIKQLPSERSKVESGFIIRLAD LERKNMPIFEDLSOTCPDLVWLQKFSAGBQVCDLRVRECEGYWLAGLGYSTKA ATIRCKETHHIILOQLTKEDVLLKKNALDEKMDLVKAIVERIYTTYARGTLKTEA GGTIKTISKEILLLSIHGSPQDOLITQLPDAANDMLCFUNSTRVNIQCALVCA LSSQNLDESSIDFDVNLGLVYIQDKENAOASASDEPKELGKFLDRLHLSVSKRL ESVLRQGLHRIALEHGNARARYDVNFVVGARLHRTKTSIFPKO"
complement (3404..4795)
/gene="CP0405"
complement (3404..4795)
/gene="CP0405"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA038250.1"
/db_xref="GI:7189330"
/translation="MLNLRVNTQIDBPRKMSNITSPIVONNSCNVYELAKNSTIIVISAILLCGALIAFLCVAAPVSYISGALLIGLIALIGLVIGIKITPMVSSKQ VFPQELVNRIRAHYKPVSPVSEAKNDLSFDLNLQHSVEGSSSTPNVSEEL OCKIDPREGIARLKNRPTASLKRLESASSRPLPSLAKTIQKVPFPLGFEISAG SKVVELHRYAKTIQGSLEBLSDTYKPEMLPTVYLIFLDRPNTSLNHLHTLVAVL TRDVFQHLRYAALNGWNLNHSDLNMTMQLLEFKHAAVQSYHLSQPSIQEDXFTNL LICIIFHRYSMQMSLKITVPADLMENLCLLDHNGRQDMEFASLIGLTYQGLIHK KESEAFSLSTLSDQFKTIRQSTNIAMFLENLATNHSSTFRSLPPIIVHPIKRSVF

gene SOPERDESSLIG" complement(4776. .5021)
/gene="CP0406" complement(4776. .5021)
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF38251.1"
/db_xref="GI:7189331"
/translation="MRPNIKKSLIFIRIRYSQSGKEQKARPEPFKSTISVILL
LEAFINENSSSIONNPNKNKINSLNRFVFKETI"
/gene="CP0407" complement(5081. .6361)
/gene="CP0407" complement(5081. .6361)
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF73665.1"
/db_xref="GI:8163426"
/translation="MRHSGCTMKSCPLTVPHIPIKNDCEHRSGLKIRITAPLLIG
LVLAVALSFPFLAAPSIAIGTALAALVILITTVALLAKSKVLPINELQKI
LYNRPKVEFFVFKTHSLVNLKLFINCMKSGTDLPPNLHKAARFGIDILKSIDLT
LEPEREILIONCPLVWLSHFIDKTESVAGEIGLKTQKVGGLPLAFHKGTYTFH
SYTRPLTILISQYKPLYSKASKNOMSPSVKTCSEIIFKELPHNMLFRDVOGISQ
FLVFFSHGITWEOQMIOLINPDKMPLCOFDKARGHCSMAFEGFNTENMEDPV
SNSVFPYVMTWQELKYLIEKVESPMKPSALVOKICVNTTHQNLKRWQVRNT
SSQNTSSLPQYAFHAKYTKLEKTISSLPINSSL"
/gene="CP0408" complement(6479. .8026)
/gene="CP0408" complement(6479. .8026)
/note="similar to GB:M28816.SP.P19568.PID:152470
GB:AJ235269, identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, ATP carrier protein"
/protein_id="AAF38252.1"
/db_xref="GI:7189332"
/translation="WTKTEBKPRGKLRSFLNPITHHEKLYLPMFLMFCITNTYTL
RDKOTLIVAGPAGSGEALPTKFWLVVPCALIEMLIVAKSNILSKOALVAVGTPF
LIFPALFPTVIVPRDVAHPPEFADRLQALIPGLGLVALIRNMTFPAFLVLELWG
SVMLSIMFMGANEITKIHBAKRFYALGIGANISLASGALVWASKLARASVSGVD
PMGISTRLLMAMTIVSGIVMASVYMKIKVLTDPFPYPMEMOKGKKGAKKMMKD
SELYIARSPYIILALIVYAGICINLEVTWKSOLKQYPMNDYSFPMNPSFTWC
VVSIVILFVGNVTRKFGMLTGLVTFVMDLTGIVFPAIVIFRNQSGVLAHFGITV
PLMAVVGAIONILSKSTKVALFDSTEMAYTIPDOQKVKGKAIDIVARFEGSG
GALIOGGLVIGSIGAMTPYLAIVILFIIAIMVLSATKMLFLAQSALKEQEVAAE
DSAPASV"
/gene="CP0409" complement(8227. .8742)
/gene="CP0409" complement(8227. .8742)
/note="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38253.1"
/db_xref="GI:7189333"
/translation="MGFMAVEQSHIKEIEKLGKAIKRVCGKNKENDLCRYLPGPSG
YMHFTLKMKSSAAPEOLLKMLKTFILSESPRTINPBRAPRGSKRRDPIHFKTD
IERVLEIARQVCKDLARFSPKRLSLKKEILISINGLIVSVMVAAYVBAKVAS
SPNLEVTSPFV"

Alignment Scores:
Pred. No.: 6,736-222 Length: 13389
Score: 2617.00 Matches: 513

Percent Similarity:	99.61%	Conservative:	0
Best Local Similarity:	99.61%	Mismatches:	2
Query Match:	99.51%	Indels:	0
DB:	1	Gaps:	0
US-09-869-433-2 (1-515) x AB002202 (1-13389)			
QY	1 MetThrLysThrguJugLysPProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20		
DB	8026 ATGACAAAACCGAAGAAAACCTTTGGAAAATTGGCCTTCTTCTGTCGCGATACAT 7967		
QY	21 ThrHisGluLeuYsLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40		
DB	7966 ACTACGAGCTAAAGAAAGTCTCGCAATGTTCCTAATGTTCTTGTGATTAACATTTAAC 7907		
QY	41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyVala 60		
DB	7906 TATACGGTGTACCGCATACAAAGACCTCTTATTTGGGAGCTCCGTGTTCTGTGCA 7847		
QY	61 GluAlaIleProPheIleLysPheThrLeuValValProCysAlaIleIlePheMetLeu 80		
DB	7846 GAGGCAATACCTTCATCAAGATTGGCTGTGTGCTCCCTGTGCTATTATCTTATGCTT 7787		
QY	81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100		
DB	7786 ATTTATGCAAGCTTAAGTAATTTTAAGTAAGCAGGCTTATTTATGCAAGTGGAGACG 7727		
QY	101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120		
DB	7726 CCGTTTAAATTTCTTGTGCGCTGTCGCGACTGAATTTATCCGCTACGCGAGTCTTTTA 7667		
QY	121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140		
DB	7666 CATCTTACAGAAATTTGCTGACCGTTTACAGGCACTCTTACCTCAAGATTTCTAGAGCTC 7607		
QY	141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyrValLeuAlaGluLeuTrpGly 160		
DB	7606 GTTGCAATCTTAAGAAACGTGACATTTGCTGCAATTTATGACTGCGAATGAGGGA 7547		
QY	161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180		
DB	7546 AGCGTCAATGCTATCTATCTTATGTTGCGGAGTTTCTATGAAATTAACAAATCACAGAA 7487		
QY	181 AlaIysArgPheTyrAlaLeuPheGlyIleGlyValAsnIleSerLeuLeuAlaSerGly 200		
DB	7486 GCAAGCGTTTCTACGCTCTTTTGATGAGCTATATATTTCTTATCTAGCTTCGTGCT 7427		
QY	201 ArgAlaIleValITrpAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTrp 220		
DB	7426 CTTGCAATGTTTGGGCTTCAAAAGTTGAGAGCTTCCGTTTCTGAAGGTGATGATCTTGG 7367		
QY	221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240		
DB	7366 GGAATTTCTTACGCTCTTTGATGAGCTATGATGATGATGAGCTGCTTCTTATGAGCC 7307		
QY	241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProAlaGlyPheTyrAsnProGlu 260		
DB	7306 AGTTACTGCTGATCAATTAAGAAAGTATGACCATCTCTCTTCTTAATCAAGAAAGAA 7247		
QY	261 MetGluLysGlyLysLysGlyValAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280		
DB	7246 ATGCAAAAGGGAAGAAAAGTCTTAACCTTAATTAATGAATGAAGATGCTTCTCTCAT 7187		
QY	281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300		
DB	7186 CTTCGTGATCTCTTATATCTTATCTTATGCTCTCTGTTATGCTTATGCTTATGCTTAT 7127		
QY	301 IleAsnLeuIleGluValIleThrTrpLysSerGluLeuLysLeuGluIleTyrProAsnMetAsn 320		
DB	7126 ATTAACCTTAATCGAAGTCACTTGAAGAAAGTCAAGCTCAATATTCCTTAATGATAT 7067		
QY	321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValIleValSerValLeuIle 340		

Db	7066	GACATATAGTAGGTTGATGGGGAACTTCCTCCCTTGGAAGTGGGATAGTATCCGTTACTATC	7007
Oy	341	MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTyrPleuThrGlyAlaLeuVal	360
Db	7006	ATGCTATTATTGGTGGGTACGCTCATTCGTAAATTGGATGGTTAACTGAGGCCCTAGTC	6947
Oy	361	ThrProValMetValLeuLeuThrGlyLysValPhePheAlaLeuValIlePheArgAsn	380
Db	6946	ACTCCTGTCACAGGTCTCCTTAACAGATACGTTTCTTCGGCTCTTGATCTTTAGAAAC	6887
Oy	381	GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal	400
Db	6886	CAAGTCTTGGGCTGGCTGCTGCTATGTCGTACAACTCCTCTCAGCTAGCTGTGGTTC	6827
Oy	401	GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys	420
Db	6826	GGAGCTATACAGAAATATCTTTCGAAATCCCAAAATACGGCTCTTGGACTCAACTAAA	6767
Oy	421	GluMetAlaTyrIleProLeuAspGlnGlnLysValLysGlyLysValAlaAlaIleAsp	440
Db	6766	GAATATGGCTATATCCCTTCCTTGACCAAGCAAAAAGTCAAAAGGATGAAGCTGCTATGAT	6707
Oy	441	ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuVal	460
Db	6706	GTAATGCGCCGCCGCTCGAAATACAGAGAGCTTTAAATCCAAAGAGTTGCTGCTT	6647
Oy	461	IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle	480
Db	6646	ATCTGTGGAAGTATTGGAGCTATAGACCCCTTATCTTGACAGATTCCTCTTTCATATT	6587
Oy	481	AlaIleTyrPleuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu	500
Db	6586	GCTATTGTTGGTGGTTCTTGCAACTAAGTTAAACAAACATATCTTAAGCCAGCTGCTGCTT	6527
Oy	501	LysGlnGlnGlnValAlaGlnGlnAspSerAlaProAlaSerSer	515
Db	6526	AAAGAACAGAGTGGCTCAAGAAATTCAGCTCTGCTTCTTCA	6482

RESULT 3
AR310754 03
WPCOMMENT

Sequence split into 13 fragments LOCUS AR310754 Accession AR310754

Fragment Name	Begin	End
AR310754_00	1	110000
AR310754_01	100001	210000
AR310754_02	200001	310000
AR310754_03	300001	410000
AR310754_04	400001	510000
AR310754_05	500001	610000
AR310754_06	600001	710000
AR310754_07	700001	810000
AR310754_08	800001	910000
AR310754_09	900001	1010000
AR310754_10	1000001	1110000
AR310754_11	1100001	1210000
AR310754_12	1200001	1300000

Alignment Scores:	
Pred. No.:	9,71e-221
Score:	2617.00
Percent Similarity:	99.61%
Best Local Similarity:	99.61%
Query Match:	99.51%
AB:	6
Length:	110000
Matches:	513
Conservative:	0
Mismatches:	2
Indels:	0
Gaps:	0

US-09-869-433-2 (1-515) X AR310754_03 (1-110000)

Qy	1	MetThrLysThnGIuLjulybProPheGjLyVysLeuAjsSerPheLeuTrpProIleHis	20
Db	104370	ATGACMAAAACCGAGMAAAACCTTTTGGMAAAATTGGCTCTTTCTTGGGCCGATACAT	1044229
Qy	21	ThrHisGIuLeuLyValLeuProMetPheLeuMetPhePheCysIleThrPheAsn	40

Db	104430	ACTCAGAGCTTAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGATTAACATTTAAC	104489
Qy	41	TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyValAspGlySerGlyAla	60
Db	104490	TATACGGTGTTAACGGATACAAAMAAACACTCTTAATGTGGGAGCTCCTCGTCTGGTGC	104549
Qy	61	GluAlaIleProPheIleLysPheTrpLeuValProCysAlaIleIlePheMetLeu	80
Db	104550	GAGGAAATACCTTTCATCAAGTGTGGCTGTGTGCCCTGTGCTATTAATCTTAATAGCTT	104609
Qy	81	IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr	100
Db	104610	ATTATATGAAAGCTAAGTAATATTTTAAGTAAGCAGCGCTTAATTTATTCAGTGGGAACG	104669
Qy	101	ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu	120
Db	104670	CCCTTTTAATTTCTTTGGCTGTGCCGACTGTAATTAATCCGCTACGCAATGTTTAA	104729
Qy	121	HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu	140
Db	104730	CATCTCAAGAAATTTGCTGACCGGTGTAAAGGCCATCTCACTCCAGAAATTTGCTAGGACCTC	104789
Qy	141	ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTProGly	160
Db	104790	GTTGCCATCTTAAGAAACTGACATTTGCTGCATTTTATGTACTTGTCTAATCTATGGGA	104849
Qy	161	SerValMetLeuSerLeuMetPheTProGlyPheAlaAsnGluIleThrLysIleHisGlu	180
Db	104850	AGCGTCATGCTATCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAAATAATCCACGAA	104909
Qy	181	AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly	200
Db	104910	GCAAGCGTTTCTACGCTCTTTTCGATCGGAGCTAATATTTCTTAATCACTTCTTGAT	104969
Qy	201	ArgAlaIleValITProLysSerLysLeuArgAlaSerValSerGlyGlyValAspProTPro	220
Db	104970	CGTGCATATGTTTGGGCTTCAAAGTTGAGAGCTTCGTTTTCGAAAGTGTAGATCCTTGG	105029
Qy	221	GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240
Db	105030	GGAAATTTCTTAACGCTTTTGAATGGCTATGACTATTTGATATCGACTTGTCTTAATGGCC	105089
Qy	241	SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu	260
Db	105090	AGTATCTGTGTGATCAATTAAGAACGTATGACCGATCCTCGCTTCTAATATCCAGAGAA	105149
Qy	261	MetGlnLysGlyLysLysGlyAlaLysProLysMetLamMetLysAspSerPheLeuTyr	280
Db	105150	ATGCAAAAGGGGAAAAAGGGCTTAACCTAAAGAAATATGAATATGCTTCTCTAT	105209
Qy	281	LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys	300
Db	105210	CTTGCTAGATCTCTCTAATATCTTTATTAAGTCTCTTGCTTATGGCTATGCTATGATTTGC	105269
Qy	301	IleAsnLeuIleGluValThrTyrLysSerGlnLeuLysLeuGlnTyrProAsnMetCAsn	320
Db	105270	ATTAACTTAATCGAAGTACTTGGAAAAAGTCAAGTGAACATCGCAATATCTTAATATGAAT	105329
Qy	321	AspTyrSerGluPheMetGlyAsnPheSerPheTProThrGlyValValSerValLeuIle	340
Db	105330	GACATATAGTGATTCATCGGGGAACCTTCTCTTGAGCTGGGGTATGATACCGTACTATATC	105389
Qy	341	MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTProLeuThrGlyAlaLeuVal	360
Db	105390	ATGCTATTTTGTGGTGTGTAAGCTATCTGTAATTTGGAATGTTAACTCGAGACCTTAGTC	105449
Qy	361	ThrProValMetValLeuLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn	380
Db	105450	ACTCTGTGCATGTTCTCTCAACAGGATATGTTTCTTCGCTGTTATCTTTAGAAAC	105509
Qy	381	GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal	400

Dd	4610	ATTATATGCAAGCTTAAGTAATATTATTAAAGTACAGGCCTTATTATTATGCACTGGGAACG	4669
Qy	101	ProPheLeuIlePhePheAlaLeuPheProThrValIleYrProLeuAAspValIleu	120
Dd	4670	CCCTTTTAAATTTCTTTGGCCCTGTCGCCAGCTGTAATTATTCGGTACAGGAGATGTTTAA	4729
Qy	121	HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu	140
Dd	4730	CATCCTACAGATTGTGGTGAACCGTTTACAGGCCATCTTACCCTCCAGGATTGGTGAAGCTC	4789
Qy	141	ValAlaIleLeuAAspAntProThrPheAlaAlaPheYrValIleuAlaGluLeuTrrGly	160
Dd	4790	GTTGGCATCTTTAAGAAACTGGACATTTGGCTGCATTTTATGTACTTGCTGAACCTAATGGGGA	4849
Qy	161	SerValMetLeuSerLeuMetPheTrrGlyPheAlaAsnGluIleThrIysIleAsn	180
Dd	4850	AGCGTCATGCTATCTCTAATGTTCTGGGGATTTTCTAATGAATTAACAAAATCACAGAA	4909
Qy	181	AlaIysArgPheYrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly	200
Dd	4910	GCAAAAGCGTTCACGCTCTTTTCGGATCGAGACTAATTTCTTACACAGCTTCGTGT	4969
Qy	201	ArgAlaIleValATrpAlaSerIysLeuArgAlaSerValSerGluGlyValaAspProTrr	220
Dd	4970	CGTGCATTTGTTTGAGGCTTCAAACTTGAAGCTTCCGTTTCTGAAGGATGATGATCCCTTGG	5029
Qy	221	GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240

Db 5030 GGAATTCTTACGCTTTTGATGGCTATGACTATTGTATCTGGACTTGTCTTATGCC 5089

100

241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGlu 260

Dh : 5080 AGTTATCTGTCGATTCAATAACAAAGCTTATTGACCCGATCCTCGCTTTCTAATCCGACGAACA E749

[illegible]

261 MetGlnLysGlyValAlaLysProLysMetAsnMetLysAsnSerPheLeuTyr 280

[illegible]

Db 5150 ATGCAAAAGGGGAAAAGGTCCTAAACCTAAATGATATGAAGATAGCTTCCCTCTAT 5209

QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300

[illegible]

DB 5210 CTGGTAGAATCTCTAAATCTTAAAGCTCTCTGGTAAATGCCAAGGAAATTC 5269

301 TLeaant.eU]eG]uValThrtmI.vSserG]nI.eU]vst.eU]G]nTyvProASnMetAsn 320

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

D5270 ATTACTTAATCGAGTGA CTTGGA AAGTCAGCTGA ACTGCA TATCTTA TATGAAT 5329

QY 321 AsPtyrSerGluPhenMetGlyAsnPheserPheTrpThrglyValValserValLeuIle 340

D5 5330 GACATATAGTGGGACTCTCTCTCTGGACTGGCGTAGATACCGTACTTAATC 5389

341 MetI,euPhaya]G]wG]vaenva]T]eArat,vspheG]vTmI,euThwG]vA]at,euVa] 360

27

Db 5390 ATGCTATTGTTGGTGGTAACGTCATTCCGTAATTGGATGTTAACTGGAGCCCTAGTC 5449

QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380

[illegible]

DB 5450 AATTCCCTGTCATGGTTCCTCCCTAACAGGATTCGTTTTCCTTCGCTCTTGTAATCTTTAGAAAC 5509

381 C]nA]aSerC]vI.aV2]n]aMatPhoC]vThvThvProI.aMatI.aV2]v2]v2] 400
 Or

[illegible]

Db 5510 C A A G C T T C T G G G C T G G T C G C T A T G T T C G G T A C A C T C C T C T C A T G C T A G C T G T G T T G T C 5569

QY 401 G I y A I a I l e G I n a s n I l e L e u S e r L y s S e r T h r I y s T y r A l a l e u P h e a s p S e r T h r L y s 420

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

5570 GGAGCTATACAGAAATCTCTCGAAATCCACAAATACGGCTCTCTTGTGACTCAACTAA 5629

431 [11Met2]aTWT]eBr0i0uB0eC]nC]uG]nI.yeVa]I.yeC]yI.yeB]aA]aT]aA00
 440

[illegible]

Db 5630 GAAATGGCCTATATCCCTCTTGACCAAGAGCAAAAAGTCAAGGTAAGGCTGCTATTGAT 5689

441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuVal 460

Dh 5690 GTAGTTCGCCCGCCGCTTCGAAATCAGAGAGCTTTAATCAACAAGATTTCGCTT 5749

Oy 461 TleCysGlySerIleGlyAlaMetThrProTyrIleuAlaValIleuLeuPheIleIle 480

Dh 5750 ATCTGTGAAGATATTGAGCTATGACCCCTATCTTGAGAGATTCTTTTATCATCTT 5809

Oy 481 AAlaIleTPluValSerAlaThrIysLeuLeuIysLeuPheLeuAlaGlySerAlaLeu 500

Dh 5810 GCTATTGCTGTTGTTCTTCGCAACTAAGTTAAACAACATTCTTATGCGCAGTCTCTT 5869

Oy 501 LysGluGluGluValAlaGluGluAspSerAlaProAlaSerSer 515

Dh 5870 AAAAGAACAGAGAGTGGCTCAAGAGATTCACTCTCTCTTCA 5914

RESULT 5

AP002546 300650 bp DNA linear BCT 25-MAY-2002

LOCUS Chlamydomophila pneumoniae J138 genomic DNA, complete sequence, section 2/4.

ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801

AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808

AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815

AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008

AP002546.2 GI:10176692

VERSION Chlamydomophila pneumoniae J138

KEYWORDS Chlamydomophila pneumoniae J138

SOURCE Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.

ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.

REFERENCE 1

AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,

Tateuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,

Matsushina,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,

Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.

Comparison of outer membrane protein genes omp and pmp in the whole

genome sequences of Chlamydia pneumoniae isolates from Japan and

the United States

J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)

20298986

PUBMED 10839753

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

2 Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,

Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.

Comparison of whole genome sequences of Chlamydia pneumoniae J138

from Japan and CML029 from USA

Nucleic Acids Res. 28 (12), 2311-2314 (2000)

20330349

PUBMED

REFERENCE

AUTHORS

TITLE

3 (bases 1 to 300650)

Shirai,M.

Direct Submission

Submitted (04-JUL-2000) Mitsunori Shirai, Yamaguchi University

School of Medicine, Department of Microbiology; 1-1-1

Minamikogushi, Ube, Yamaguchi 755-8505, Japan

E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,

Fax:81-836-22-2415)

On or before Sep 15, 2000 this sequence version replaced

gi:61722290, gi:61722292, gi:61722294, gi:61722296, gi:61722326,

gi:6172338, gi:6172330, gi:6172332, gi:6172334, gi:6172336,

gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,

gi:6172390, gi:6172392, gi:6172394, gi:6172396, gi:6174668,

gi:6655174, gi:6655176, gi:6655178, gi:6655180, gi:8547433,

gi:8547438, gi:8978640

Location/Qualifiers

1. .300650

/organism="Chlamydomophila pneumoniae J138"

/mol_type="Genomic DNA"

/strain="J138"

/db_xref="taxon:138677"

/note="synonym:Chlamydia pneumoniae (strain J138)"

144. .935

/gene="CpJ0267"

144. .935

/gene="CpJ0267"

gene

CDS

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BA08477.1"

/db_xref="GI:8978641"

/translation="MSIMSLINKTNALINQEPAPVCUNAMPKYNIDNRKTPACTVTL

VITLMLTIGTVILVLMSPGSLVSTIGSVTLTALFIVGLIKSLAM

OYKROEVVKORKEPSPRIKPNNDVHLTSLSPIDIESPSPASTPVSKRIACSG

VALVAGTLLIGAVSVFPGTGLQALCTGPGCTGALPVGAGARTSLIAGIM

YLVTLYLSALSERNKTVDKRNENITVLTRECRQREKALKE"

957. .1343

/gene="CpJ0268"

957. .1343

/gene="CpJ0268"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BA098478.1"

/db_xref="GI:8978642"

/translation="MSQCSSSTSTWEMKSPVPMNKMPPTSPSPSEDEFILAYEP

FVLPKTPDERAKANPPTSPVNGTDLNPLGQNEQNNANPOTSSNPTSLPA

PERLPETBENSQSEBQSQNNEDLIG"

complement (1495. .2472)

/gene="CpJ0269"

complement (1495. .2472)

/gene="CpJ0269"

/codon_start=1

/transl_table=11

/product="dipeptidase"

/protein_id="BA098479.1"

/db_xref="GI:8978643"

/translation="MTIDMKCDLLSHPRCKDPAVRCSPKQSLSGSVROVCAIFP

HSKGERNDKQNSLFSLSNPQYDGLSLSTREBENGSSSKSLSTRISFMSALGD

DTPLGLTLKLTILTRQGLAYLTGWKNDNFGGTGTEPKLSNDKGLDMLMEL

GVIDLSHCSLKAEDILDTADKLPLAVIASNSFRSLDVRNLDVAHAETVVR

KGVIGNLVRSYVGDSDIGLEKVLHAENIGLISVLGDFPFAEDENFPFNESCS

AEAPVNLQNLIRHIFSKRAESILSRRAEKLQVIVEQVNPKITDKL"

complement (2487. .3347)

/gene="ywlC"

complement (2487. .3347)

/gene="ywlC"

/codon_start=1

/transl_table=11

/product="Sua5 superfamily-related protein"

/protein_id="BA098480.1"

/db_xref="GI:8978644"

/translation="MPDKKAQITFSLPVNSAIHQKIVALPDTVYGFVLSLYASEA

EEELVADKREPSKAPALVYNSIEDENISGYPLSPAKLAQLPFGAIVLVKHNRP

RFPKETLAIRFIVDSVREIVDHGCTLIGTSANLSEPSKALTQEIFADPADDLICF

DGPCSHGSESTVAVSDPIVYREGLSRSVTENIAGTEAKIFRTSHAFSKHITKV

KNOEOVSPISGSLDKGVCEHPKRNFTRLREALKKKTPSIVFTYDNTSDYBEL

FPFLSPYYIE"

3656. .4387

/gene="CpJ0271"

3656. .4387

/gene="CpJ0271"

/codon_start=1

/transl_table=11

/product="lysophospholipase esterase"

/protein_id="BA098481.1"

/db_xref="GI:8978645"

/translation="MTDVSFFRRKIKNTEAIEGRGNPDPIILCHGVGSLADNLTFP

PSICFSKLRPTIIPNGIILPLENDRGRSACPLAVNLLOELSRILYANGVNLQERY

DELFDVLETPKALELILNLNRPYNEIITIGFSOGALILATLVTSQNPVAGALF

AGARLNPQMEBGLKQCAQVPLQSHGDEBELPYHLGAHLNDLLLTKNQGVSPFG

GHEIPSVFOKQVTVPMNIDPARG"

complement (4359. .5255)

/gene="dnaX_2"

complement (4359. .5255)

/gene="dnaX_2"

/codon_start=1

/transl_table=11

/product="DNA polymerase III gamma and tau"

gene

CDS

```

gene
CDS
    /protein_id="BA98482.1"
    /db_xref="GI:8978646"
    /translation="MHLFEENOGWEALLRVYHOEVPAILLIGFTLIPVQDKABOLA
    SEILLSSPSSSEKHSOKTHPDYQFPECKGRHSLDLPGRIGKQIYIPFPANRKT
    YIIHEADRTMLAAISAFKVEEPPKIAVILITAKVQRLPKTIISRLSILFLERGK
    ILICSGEYSLFRPAOCEIPVEEQSIKSSSECKVLDKQRFKORFESLIEFLERDRY
    TLNLGLASALANYEHEVKIILQLPLLDKVLILVESACSLNNSSSAASVLEVAIO
    LVSLOYEKELVSPGQDLN"
    complement(5246..5866)
    /gene="vdk"
    /complement(5246..5866)
    /gene="vdk"
    /codon_start=1
    /product="thymidylate kinase"
    /protein_id="BA98483.1"
    /db_xref="GI:8978647"
    /translation="MPVIEGEGSGSKSLAKALGQVADRKVLLTREPGLIGE
    RLBDLLEPHLEISRCCEFLPLGSPAGHIOEVIIIPALDGVVICERHDSITVYQ
    GIAEGADPVDLCSKVGPPTLPPLPVLIDIPADIGLGRKHROKFPFKPKPLS
    YHNRIRBGFSLIASADPSRYLVDARSLASLIDKWLHTQGLCT"
    complement(5871..8375)
    /gene="gyra_1"
    /complement(5871..8375)
    /gene="gyra_1"
    /codon_start=1
    /product="DNA gyrase subunit A"
    /protein_id="BA98484.1"
    /db_xref="GI:8978648"
    /translation="MFKKDEIIVKRNLEEMKESYLRESVSIISRALPDIRDLKLS
    ORBVLAMKOLSLSPGKHKRCAKIGDSTGDYHPHESISYPLVPMANMRYPLS
    VDOGNFGSIDDPRAARTYERLTLTAATLMEDDKQVVDVVPYDETRKHPVVP
    SKPRLNCGSSGLAVGMATNIPPHNLGELIEATLLAPASVBDILVMPGRLD
    TGGIICGSEGIIRSTYTGKIKIVRAHVENEDKRESIIITEMFYNNKQRLIQ
    IANLVNKEKTLAGISDVDESDDKGIIVLEIKKGESEIIINLYKFTDQVFGAM
    LAIDNLPRTMSIHRMISAWIRHREKIVIRRTREYELKATRAHVEGLYKALSCDA
    LVTTIRSGNKEHAKERIIRSGFTEPQALAILLELRYOGLAEAEIKOEVSEILNK
    IAYKQVLSDEGLVKDIIIRNELODLKHVYARTTTEFPADIDRIEDLITNESYI
    TISGDDYKRPVAVKEQRGGHGVTFPMKKGAGFLKAVSAFTKDYLLITNFEQ
    CYWLKVMQLEPGRERAKRPIINLEBGRIGEBELALANTKNDNAGFLFLAKRGV
    KKYSLDAFNPBRKGIIRALEIDEGDEILAACHIVSDEKWLFTHLGMAVRFHEKVR
    PMGRITAGVGVGLKNEDEKVSQOIVTENOSVLIYCDQGFGRSLVEDPRTNRGV
    GVSILINENGVGLAIPLYDHSIILMSOGQAIRINODVAVKRSIQGVRLVHL
    KEDDALVMEKSSNENDVLSGSEECSDTVSLR"
    complement(8391..10808)
    /gene="gyrB_1"
    /complement(8391..10808)
    /gene="gyrB_1"
    /codon_start=1
    /product="DNA gyrase subunit B"
    /protein_id="BA98485.1"
    /db_xref="GI:8978649"
    /translation="MDPKENYDASATVTEGLQAVRPPGMYIGDTGTLHLIYVE
    VVNSIDEAMAGYCSRIDVRLIEDGIVINDGRGRIIVYHRESAKOGSEVALYVE
    LTYLHAGKFDKDSYKSGGLHGVSCVNAISEKVATYFKKCTOMFSGSIPPT
    PLQVSVSDQGEIIVFPDKIFSTCTPFRSILMKRLAELANRITTLVREDDRV
    SPDKVTFEYEGQIQSFVSYLONKESLFPSPRIYICGRVDDDEIEFEALQNNNGS
    ELVSYANRNIPTROGTHLTGFTALRVNNTYIKANLKNKMLATGMDIREGLVA
    VISVKNPQFEGOTKOKLNSDVSVAOVGVALIIPRENPOLARMIVDVAVPA
    OAREAAKARELTAKSALDSARLPGLIDLCRDKPKCMYIVEGDSAGSAGKQGD
    RRFOALIPRIGKILNVEKARLOKIPONOEITTIATLIGCIGIDNFKLIRRTIIT
    MTDADVDSHIRTLLTFPRYHMTALINENCVYIAOPLVSKKQOFRIILSEKED
    SYLLMGNTSSSILFKSTERELGEALLESINVLDESFINLLEKALIPFSEFLMY
    KEGIGLYLAIATPGQGRVLYVSDKEEALAOEETHFKIIEIKVAVPVDIQWQ
    LKEGMDLISYLLPQKNEIYVGNDSFSCVSCYTLSEVINYLNKGRKGIETORVKG
    LGEAMDOLDMTTMMNPRQRTLIHVSLEKDAYEADHITFMNGEVPFRPRETIESHALSI
    RINNDI"
    complement(10812..11159)

```

gene
Alignment Scores:
Pred. No.:

1.45e-219 Length: 300650

Score:	2610.00	Matches:	512
Percent Similarity:	99.42%	Conservative:	0
Best Local Similarity:	99.42%	Mismatches:	3
Query Match:	99.24%	Indels:	0
DB:	1	Gaps:	0
US-09-869-433-2 (1-515) x AP002546 (1-300650)			
Qy	1	MetThrLysThrGluGluValProPheGlyLysLeuArgSerPheLeuTyrProIleHis	20
Dd	93996	ATGACAAAACCGAAGAAAACCTTTGGAAAATGGCTCTTCTTGTGGCCGATCAT	94005
Qy	21	ThrHisGluLeuLysLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn	40
Dd	94056	ACTCAGAGCTTAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTGTATTAACCTTTAAC	94115
Qy	41	TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla	60
Dd	94116	TATACGGTGTACCGCATCAAAAAGACACTTATTGTGGAGCTCGTGGTTCGTGCA	94175
Qy	61	GluAlaIleProPheIleLysPheMetPheValValProCysAlaIleIlePheMetLeu	80
Dd	94176	GAGGCAATACCTTTCATCAAGTTCCTTGTGTGCTGCTGCTATTAATCTTATCTT	94235
Qy	81	IleTyrAlaLysLeuSerAsnIleLeuSerLysGluAlaLeuPheTyrAlaValGlyThr	100
Dd	94236	ATTTATGCAAGCTTAAGTAAATATTTTATGACAGGCTTATTTTATGACAGGAGACG	94295
Qy	101	ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu	120
Dd	94296	CCCTTTTATTTCTTCTTGGCCGTCGTCGACGTGATTTATTCGTCAGGAGTCTTTA	94355
Qy	121	HisProThrGluPheAlaAspArgLeuGluAlaIleLeuProProGlyLeuGlyLeu	140
Dd	94356	CATCTCAACAATTTGCTGACCTTTTACAGGCACTCTTACCTCAGATTTGCTAGACTC	94415
Qy	141	ValAlaIleLeuArgAsnTyrPhePheAlaIlePheTyrValLeuIleGluLeuTyrGly	160
Dd	94416	GATGCCATCTTAAGAAACCTGACATTTGCTGACATTTTATTTATCTTGTGAACTAGGGA	94475
Qy	161	SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIleThrLysIleHisGlu	180
Dd	94476	AGGTCATGCTATCTCTTAAGTCTTGGGATTTGCTGATTAATAAATTCAGAAATCAGAA	94535
Qy	181	AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly	200
Dd	94536	GCAAGCGTTTCAACGCTCTTTCGATCGAGCTAAATATTTCTTACTACGATCTTCGT	94595
Qy	201	ArgAlaIleValIleTyrAlaSerLysLeuArgAlaSerValSerGluGlyValAspProTyr	220
Dd	94596	CGTGCATTTGTTGGCTTCAAAAGTTGAGGCTTCGTTCTGAAAGGTGAGATCTTGG	94655
Qy	221	GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240
Dd	94656	GGATTTCTTCTTACGCTCTTTGAGGGATGACATTTGTAATCTGGAATCTTCTTATGGCC	94715
Qy	241	SerTyrTyrTrpTyrIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGlu	260
Dd	94716	AGTTACTGGTGATCAATAAGACGATTAAGCCGATCTCTTCTTAATTCAGAAAGAA	94775
Qy	261	MetGluLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr	280
Dd	94776	ATGCAAAAGGGGAAAAGGTCCTAAACCTAAATGAATGAAGATGATCTTCTCTAT	94835
Qy	281	IleAsnAspSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys	300
Dd	94836	CTTGCAATGATCTCTTATATCTTATTAATGCTCTCTGTTATGCTTAATGATTTGC	94895
Qy	301	IleAsnLeuIleGluValIleThrLysSerGluLeuLysLeuGluIleTyrProAsnMetAsn	320
Dd	94896	ATTAACCTTAATCAGAAAGCTTGAAGAAAGTCAAGTCAATGCAATATCTTAATATGAAAT	94955
Qy	321	AspTyrSerGluPheMetGlyAsnPheSerPheTyrThrGlyValValSerValLeuIle	340

QY 201 ArgAlaIleValIleThrAlaSerIleuArgAlaSerValSerGluGlyValAspProTrp 220
 DB 188095 CTTCTATTATCTGGGCTTCTTAAGCTACTGCTAGCGCTACAGGAAATACATCTCTGG 188036
 QY 221 GlyIleSerIleuArgIleuIleuMetAlaMetThrIleValSerGlyIleuValIleuMetAla 240
 DB 188035 GGACTTCCCTTACCTCTTAATGAGTATGTTATCTTCCGGCGGCTAATTTATTCTTA 187976
 QY 241 SerTyrTrpTrpIleAsnIleuValIleuThrAspProArgPheTyrAsnProGluGlu 260
 DB 187975 TGCTACTGCTGATGAAATGAGCGCTTCTTACAGATCTCAAGATCTTATGATCTTAAGAA 187916
 QY 261 MetGlnIleGlyIleGlyIleValIleuProIleuMetIleuAspSerPheIleuTyr 280
 DB 187915 TTGAACAAATCCAGAAA---TCCAAGCCAAATAATGAGATAGAGAAAAGCTTCTTAT 187859
 QY 281 LeuAspArgSerProTyrIleIleuIleuThrLeuIleuValIleuAlaTyrGlyIleCys 300
 DB 187858 TTAGCAAGATCTCTTACATGTTATTTACTAGCCCTTCTTATTTGCTACCGCGCTTGT 187799
 QY 301 IleAsnIleIleGluValIleThrIlePheSerGlnIleuIleuGluTyrProIleMetAsn 320
 DB 187798 ATTAACCTCGTAGAAGTTAACCGAAGACCAATTGAAATGCATACCAATCTCTAAC 187739
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTyrThrGlyValIleValSerValIleu 340
 DB 187738 GAATATACCGACTTATTTGGTAACTTCTCATTTCTGACCGGGGCTTATCTCTTATCTGG 187679
 QY 341 MetLeuPheValIleGlyIleAsnValIleArgIlePheGlyTyrIleuThrGlyValIleuVal 360
 DB 187678 ATGTTATTTGTCGGTGTACCTTATCCGTAATTCGGTGTGTAAACAGAGCTCTCGTA 187619
 QY 361 ThrProValMetValIleuIleuThrGlyIleValIlePhePheAlaIleuValIlePheArgAsn 380
 DB 187618 ACTCCAGTTATGCTATGTTAAACAGCGCTTCTTCTTCTGCTCTGATTAATCTTCCAGAGAT 187559
 QY 381 GlnAlaSerGlyIleuValAlaIleMetPheGlyThrThrProIleMetLeuAlaValIleVal 400
 DB 187558 CAGGCTTCTGCAATCTGACTACGTTAGGAAACACACCATTAATCTTCTGCTTATGTT 187499
 QY 401 GlyAlaIleGlnAsnIleuSerIleuSerThrIleValIleuPheAspSerThrIleVal 420
 DB 187498 GGGGCTATCCAAACATCTCTAATCTCAATCCAGAAATACGCCCTCTGATCCAAACAAA 187439
 QY 421 GluMetAlaTyrIlePheIleuAspGlnIleuGluIleuValIleGluIleuAlaIleuPhe 440
 DB 187438 GAAATGGCTTACATCTTCATGATCAAGAAACAAAGTTAAAGAAAGCTGCTATTTGAC 187379
 QY 441 ValValAlaAlaIleArgPheGlyIleuSerGlyIleValIleuIleGlnIleuValIleuVal 460
 DB 187378 GTAGTTGCTGCTGCTTTCGAAAAATCCGGCGGCTTCTTATTCACAAAGTCTTCTAGTG 187319
 QY 461 IleCysGlySerIleGlyAlaIleMetThrProTyrIleuAlaIleIleuIleuPheIleIle 480
 DB 187318 GTTGTGGAAGATTTGAGCGAGTGAATCTTACCTAGCAATGACTCTTCTTGTGATTAAT 187259
 QY 481 AlaIleIlePheValSerAlaIleThrIleuAsnIleuIleuPheIleuAlaIleGlnSerIleu 500
 DB 187258 GCTGATGCTTATGCTGCTGCTCAACAACTTAATTAATCTTCAATCTGCGATC 187199
 QY 501 LysGluGlnIleuValAlaIleGlnIleuAspSerAlaProAlaSer 514
 DB 187198 AAAGAACAAAGACTTCAGAGAGAAAGACCGCTGCTCTTCT 187157

RESULT 7
 AB001281/c
 LOCUS 14168 bp DNA linear BCT 30-OCT-2000
 DEFINITION Chlamydia trachomatis section 8 of 87 of the complete genome.
 ACCESSION AB001281 AB001273
 VERSION AB001281.1 GI:3328454
 KEYWORDS
 SOURCE Chlamydia trachomatis
 ORGANISM Chlamydia trachomatis

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 FEATURES
 SOURCE
 Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 1 (bases 1 to 14168)
 Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,
 Koonin, E.V., and Davis, R.W.
 Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis
 Science 282 (5389), 754-759 (1998)
 99000809
 9784136
 2 (bases 1 to 14168)
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
 Olinger, L., Grimwood, J., Davis, R.W., and Stephens, R.S.
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 Nat. Genet. 21 (4), 385-389 (1999)
 99206606
 10192388
 3 (bases 1 to 14168)
 Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,
 Koonin, E.V., and Davis, R.W.
 Direct Submission
 Submitted (20-May-1998) Program in Infectious Diseases, University
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
 Location/Qualifiers
 1..14168
 /organism="Chlamydia trachomatis"
 /mol_type="genomic DNA"
 /strain="D/UW-3/CX"
 /db_xref="taxon:813"
 86..847
 /gene="flaA"
 /note="synonym: CT061"
 86..847
 /gene="flaA"
 /codon_start=1
 /transl_table=1
 /product="Sigma-28/WhiG Family"
 /protein_id="AAC67652.1"
 /db_xref="GI:3328455"
 /transl_table=1
 /product="WKTHDLDVTDWQIYWSYKEITHRDVLESIYPLVKNVAHRLASGH
 PSHVMEKDLVAGVEGLIRAVERPDPKSKRFESYALFIKAIIIDLRKQDWPBSV
 YORANRLADAMDLSRQTLGKPEPTGELCEYLNISQDLSHWFSRSSRALVLSDDPS
 CODDDEGLAEERLADERAETGYDVIRKKAISLTETALALDEKQVVALYYDDL
 VKIKIKILCVSSSRVSQHSKLLKLRGLTSLSL"
 1006..2244
 /gene="tyrS"
 /note="synonym: CT062"
 1006..2244
 /gene="tyrS"
 /codon_start=1
 /transl_table=1
 /product="tyrosyl tRNA Synthetase"
 /protein_id="AAC67653.1"
 /db_xref="GI:3328456"
 /transl_table=1
 /product="MOOLIDNLKRGILINDNSAGLESLTYVPVSAVLGFDPTAPSLHIG
 HMGICFRLRLAAGITPVALVGATGATGIDPSSGVSERSLIDQALVNSKKAAL
 ASYIPGIRIVNADMLGSLSMVDPTRVUGGFRGSLADLVYKORVYSESGISYTF
 SYLLQSYDFAHLPEKHNVLVQCGSDQMGVITSGIDYIRRGIGQAYGLTYPLTTS
 KGKKIGTGESGTLWLDPALTPPELFPYFRLPQELSKWRTLTLDNEIFALDER
 LTPSPQAKYVIAEVIKVDVHSGELQAQAAATESFPASGKSTTEBELVALVBSGV
 VKVARADLIGKRMIDIVELGFCSSRGQARRLIQORGLYINQELADEQSLDGTQC
 FDRVVLISGKRRKKQVIDLN"
 2254..3696
 /gene="gnd"
 /note="synonym: CT063"
 2254..3696
 /gene="gnd"
 /codon_start=1
 /transl_table=1
 /product="6-Phosphogluconate Dehydrogenase"
 /protein_id="AAC67654.1"
 /db_xref="GI:3328457"


```

/translacion="MAPNTDIGLIGLAWGNKLVLMNVHGFVSIVYRNPAKTEETL
KDHGSGALGFTTIOBFVOSIKRPRKIMIKRAGPVDMDIAILPELEGILLD
GNSYILDSBORYVDLKEGILPVMGISGSEBGRKPSIMPGSNTIDAMPALIPROS
IAQVDRPCCSMIGTGAGHFWAVANGIBYGIQICETYEILKTRILNLSJQCN
IFPEWNTDNTSYLIGAAAVLAKDENGNAISTIDVAGOKGTGWADIAKAI
PMSLITSVLARVYSTWKEVTKAOFPGIPLICOPPOBASAIEDVRALVAAKI
SYAGFMYLLKOVSDKMDLIGELALIMGCGTIOGAFIDKIHQGENPEHSLIL
ODYFKYLLPDSGTFRRARVLAIGSGVALICLSALSFDYGRYVDSLFLVQGLRY
FGAHGYRDCPRSEFYHTDMLTKTFRY"
complement (3757. .5565)
/gene="1epa"
/note="synonym: CT064"
complement (3757. .5565)
/gene="1epa"
/codon_start=1
/transl_table=11
/product="GTPase"
/protein_id="AAC67655.1"
/db_xref="GI:3328458"
/translacion="MKRYKIENTRPSIINHDKSTIADRLLESTSTEQEMREO
LBSMDLEREGITTKAPVMTYEYGETBELIDPGHDSIEVSRLACBGA
LLIVDAQVQVQSLANVYLALERDEIIPVANKIDLPAAQPEAIKQIEBEFLDTS
NTIACSAKTQGIPEILESIRLVPPKPEQETELKALIFDSHYDPVYGIMVYVRS
GEIKKDRITFMAKSGSFEVLGIGAFLEPATMEGSLRAGOVGYFIANKKYDVKI
GDVYTVKHPAKEPLEGGEKEIPVPAIGVYIDSDPDTKDALGRLQNDALTBEO
ENHSLGGRRCGFLGHLIETPERISREDDIITAPSVIYKVLNKGKTLFISY
PTAYPDALIBHEEPVYNIITPQSTINSLSLCHKGICLKTMDLQHLVISTY
ELPLNEIVSDPNDKLSVTGYSDFRLDQYKGAIIKLEIILNDEAVAFSLVHR
DKESKRSICEKLVADIPOLEFKIPIOAINKKIIRARETIRALAKNVYAKCYGDDIT
RKRLMDKQKQKKEKKEFGKVSIPNTAFVEVLKME"
complement (5751. .7337)
/gene="CT065"
complement (5751. .7337)
/gene="CT065"
/codon_start=1
/transl_table=11
/product="ADP/ATP Translocase"
/protein_id="AAC67656.1"
/db_xref="GI:3328459"
/translacion="MTQYAEKPFGLKSLFIMPHMELKVLPMFLMFCISFNYTIL
RDTDTLIVAPSGGAELIPKILMLVPSVVMILYAKLSNLTNQAIFVALSPF
VVFALPFPVYIPCRHILHPTAFADTLOSILSGPMGFIAMLRMTFAVYVSELG
SVMLSMFGMPANEITKISEAKRYALFGVGNALLISGPATIMWSKILASIGEDV
PMGVSLEYLMAPECSAIIAACCYMMNRVYUDPRYPNPAELKAKSKSKMSGSEF
SYLRSYMLLALVITCYGICMLVVTMSQKQFENPNDYSAFMGPFSEVTV
SVFVPLFIGNVIRRFGLTGVLPIMVLTGAFVAVI FRDHATGLVAAAGTTP
MLAVVGAIONIISSTKYALFDATKEMAVIPLDOEQKVGKAAIDVVAARFGSGGS
LIQOGLVVGSGISGAMPPLAVALFIMVLTATLNTKFLAASAKQELAEAAA
AEKASSAKESAPALIEGVS"
complement (7688. .8164)
/gene="CT066"
complement (7688. .8164)
/gene="CT066"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67657.1"
/db_xref="GI:3328460"
/translacion="MATAOITIOEIEBOLITKAISKVSGSKENDLCRYLPPTGYWH
HFLTKKMSAPBOLLKMLKTFVIDSTPSISPKPAPGSKRGRDFNVTKDIR
VLELRQVSKDLARSPKKPLPSLRBELIRIRNNIVETELMNAVYELSQTE"
8822. .9802
/gene="YtgA"
/note="synonym: CT067"
8822. .9802
/gene="YtgA"
/codon_start=1
/transl_table=11
/product="Solute Protein Binding Family"
/protein_id="AAC67658.1"
/db_xref="GI:3328461"
/translacion="MSFFHTRKYLLRGLCLAGCFIAMSSCSSRGNOPAESIVYL
SNRMICDVCRIIGDVRVKNIVLIDGAIIDPHSYEMVAGKDDBRAMSQLITCNGIGLEH

```

```

SASLRKHEGNPKVVDLGGRLINKNCFDILSEEGFPDPPIIWTDMRWGAAVKEMAAL
IOQFPVEEDFOKNADOIILSEMELEDRMARISISTPENKRYLVTHGNAFSYTRRYL
SSDAERVSSEMRKCSIPSGSLSPRAOISIRDIRVVEYISANDVEVPELDTINQAL
RKIVSCSGGQKRIKLKSPLYSDNVCDNFSTFOHNVRIITBELGGTILE"
9795. .10574
/gene="YtgB_1"
/note="synonym: CT068"
9795. .10574
/gene="YtgB_1"
/codon_start=1
/transl_table=11
/product="rRNA methylase"
/protein_id="AAC67659.1"
/db_xref="GI:3328462"
/translacion="WRDNALAMSVEDLCVNYHSDVLCITHTSLPAGMAALITPGV
AGKSTLLKASLGLIRASSGQSLFPGQFBSKVHRIAYMPORASVMDPFTVLDVLM
GCYCYGKGINWRIISTDRCAMRLIERVGEAEANRAGIIGKSGCGOORAPFARSLMOXA
DLYIMDELPSAIDMASYOMVDVLOELKEGKTIYVIHDLISVRLKFDHVIILNKL
VCSGSVECLTKRAIFQAYGCELELDYTLKLSRGKYGQSC"
10575. .11930
/gene="YtgC"
/note="synonym: CT069"
10575. .11930
/gene="YtgC"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAC67660.1"
/db_xref="GI:3328463"
/translacion="MLSCIFQDTRIFLSSFLAVSLICMTTALMGITLILVEROPLSESL
SHACYGLIGALLSYKVPFSDSLVWIIIFPGCLAVLGLGISPEKRAMKQDLSL
CLVLSFPFGVGLVSVVDCPLVKNKINAYLGOAATLGYTEAKALIIPLCLSVY
LMMWYROIISATIDREFPASCGRIRTAELVYVFSIYVSVRSVGLIILISMPVA
PULSARQLSRSLTILILSIFGIGLGCYFVAFPCQTVGKPISTILPTGPLY
VFPAGVLFCLIFSWKGTWITFRPKKFLFSRDEHLLKIFMYLREQNTYQVMD
FVRSRYOEYFGDKVFPFRMFLCKKGLVSCSEHMGSLTDKGLADAAKLVRARLME
SVYSLQDLDFKQNEVHHFABEMEHVLTDELDSITLSQMLQDPDVPQREIIPKTRKSDG
C"
11920. .12876
/gene="YtgD"
/note="synonym: CT070"
11920. .12876
/gene="YtgD"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAC67661.1"

```

Alignment Scores:

Pred. No.:	4.12e-180	Length:	14168
Score:	2147.00	Matches:	410
Percent Similarity:	88.00%	Conservative:	52
Best Local Similarity:	78.10%	Mismatches:	49
Query Match:	81.63%	Indels:	14
DB:	1	Gaps:	3

US-09-869-433-2 (1-515) x AB001281 (1-14168)

Qy	1	MetThrLySthrgLugLyuSProphegLYLeuAArgSerPheLeuTrpColIeHis 20
Db	7337	ATGACTCAACCGCGGAAAAACCTTTTGGAAATGGCGCTCTTTCCCTTGCGCGATACAC 7278
Qy	21	Th-HiGluLeuLySValLeuPromeT-PheLeuMet-PheCysIleThr-PheAsn 40
Db	7277	ATGCATGAGCTGAAGAAAGTTCTGCCAATGTTCTTAATGTTCTGTGATTCATTCAT 7218
Qy	41	Tyr-Thr-ValLeuArgAspThrLySAspThrLeuIleValGlyLysProGlySerGlyAla 60
Db	7217	TACACGATTTTGAGAGTACAAAGACTCTTATGCTTACACACCGGAGATCGGAGCA 7158
Qy	61	GluAlaIleProPheIleLySPhetPLeuValValProCysAlaIleIlePheMetLeu 80
Db	7157	GAGGCAATTCCTTCAATTAAGTGTGCTAGTGTGCTCCTTCTGCTGTGTGCTTCAAGCTT 7098

QY 81 ILeTyrAlaIysLeuSerAsnIleuSerIysGlnAlaIeuPheTyrAlaValGlyThr 100
 DB 7097 ATCTACGCCAAGCAATGATATTTTGAACAAAGCGCTCTTTCTGGCAGTGTCTTCA 7038
 QY 101 ProPheLeuIlePhePheAlaIeuPheProThrValIleTyrProLeuArgAspValIleu 120
 DB 7037 CCATTCCGTTGATCTTCTGCACTGTTCCCTGCGGATCACTACCTTGCCGTCATATCTT 6978
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleuProProGlyLeuLeuGlyLeu 140
 DB 6977 CACCCACAGCGCTTCGCTGATCACTCAACATCTCTCCCTCCGGATTTATGGAATTC 6918
 QY 141 ValAlaIleLeuArgAsnIlePhePheAlaIlePheTyrValIleuAlaGluLeuTrpGly 160
 DB 6917 ATTGGAATGCTACGCACTGACATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6858
 QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaGlnGluIleThrIleIleHisGlu 180
 DB 6857 AGCGTTATGCTCTCTTATGATGTTCTGGGCTTGGCAATGAATTAATTAATTAAGCAA 6798
 QY 181 AlaIysArgPheTyrAlaIeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 DB 6797 GCTAAACGCTTCTACGCTCTTTCCGAGTAGAGCTAACGATCTCTTAATTTCTGCT 6738
 QY 201 ArgAlaIleValTrpAlaSerIleuArgAlaSerValSerGluGlyValAspProThr 220
 DB 6737 CCAGCTATTATCTGCTCTTCTTAATTAACGCTGCTGCGAAAGGCGTGAACCAATG 6678
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleAlaSerGlyLeuValLeuMetAla 240
 DB 6677 GGAGTCACCTCTATCTTCTTAATGCTATGTTCTGTTCTTGGCATATATGCTGCT 6618
 QY 241 SerTyrTrpTrpIleAsnIleAsnValIleuThrAspProArgPheTyrAsnProGluGly 260
 DB 6617 TGCTACTGCTGATCAACCGCTACCTGCTTACAGATCTCAATCTCAACATCTGCAAG 6558
 QY 261 MetGlnIysGlyValIysGlyAlaIysProIysMetLeuIysAspSerPheLeuTyr 280
 DB 6557 CTC--AAAGCTAAAGAA--CTTAACCTTAAGATGACAGAGGCGAAAGCTCTCTAT 6504
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 DB 6503 CTGTTAAGATCTCTTATGCTTCTTCTGCTCTTCTGATTAATGCTACGGAATCTGC 6444
 QY 301 IleAsnLeuIleGluValIleThrIlySerGlnLeuIlyLeuGlnTyrProAsnMetAsn 320
 DB 6443 ATTAACCTCTGTAAGTAACCTGGAAGCAACCAATCAAAATCCCTTAATCCAAAT 6384
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpTrpGlyValIleSerValLeuIle 340
 DB 6383 GACTACACGCGATTTATGGGAACCTCTCTTCTGGAAGAGATGATCTGATTTGTA 6334
 QY 341 MetLeuPheValGlyIysAsnValIleArgIysPheGlyTyrPLeuThrGlyAlaLeuVal 360
 DB 6333 ATGCTCTTCACTGGGATGTAACGTTATCGGTATCGGTGTTAATCGAGCTTTGTT 6264
 QY 361 ThrProValMetValLeuLeuThrGlyIleValIlePhePheAlaLeuValIlePheArgAsn 380
 DB 6263 ACACCTATCACTGTTCTGTAACAGAGCTCTCTTCTGCTCTGCTATCTCAGAGAT 6204
 QY 381 GlnIleAspGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValIleVal 400
 DB 6203 CATGCTACAGGTTTATGTTGCTGCTTGGGAACAAACCACTTAATGCTGCGTGTGTC 6144
 QY 401 GlyAlaIleGlnAsnIleLeuSerIysSerThrIlyTyrAlaLeuPheAspSerThrIys 420
 DB 6143 GGAACCATTAATAATATCTTCCAAATCACTAATATGCTCTTGTGACCAACTAA 6084
 QY 421 GluMetAlaTyrIleProLeuAspGlnGlnIlyValIysGlyValAlaIleAsp 440
 DB 6083 GAGATGCTTACATCCCATGGAACCAAGCAAAAGTTAAAGGGAAGGCTATTTGAC 6024

QY 441 ValValAlaIleArgPheGlyIysSerGlyIleAlaLeuIleGlnGlnIlyLeuLeuVal 460
 DB 6023 GTTGTCCCGCTACATTCGGTAAATCTGGGGGATCTTGTATCCAAACAGATCTTGTGTC 5964
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuPheIleIle 480
 DB 5963 GTTGGCGAAGCATCGCGCTATGACTCATTCCTACTGTTGCTCTCTTGGCAATATATC 5904
 QY 481 AlaIleThrLeuValSerAlaThrIlyLeuAsnIlyIlePheLeuAlaGlnSerAlaLeu 500
 DB 5903 ATGCTCTGTTGACATCTGCACTAACTAAACAACTATGTTGCTGCAATCGCTGCT 5844
 QY 501 LysGlnGlnIlyValAlaGln-----Glu 508
 DB 5843 AAAGAACGAAATTAAGAGAAAGCTGCAACAGACAGAGAAAGAGCTTCTTGGCGCTAA 5784
 QY 509 AspSerAlaProAla 513
 DB 5783 GAATCCGCTCTGCT 5769
 RESULT 8
 TAU10586
 LOCUS TAU10586 1587 bp DNA linear BCT 25-FEB-2003
 DEFINITION Chlamydia trachomatis np1 gene.
 ACCESSION AJ010586
 VERSION AJ010586.1 GI:4538859
 KEYWORDS NP1 gene; nucleoside triphosphate transport protein 1.
 SOURCE Chlamydia trachomatis
 ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE
 1 Tjaden, J., Winkler, H.H., Schwoppe, C., Van Der Laan, M., Mohlmann, T. and Neuhaus, H.E. Two nucleotide transport proteins in Chlamydia trachomatis, one for ne nucleoside triphosphate uptake and the other for transport of energy
 JOURNAL J. Bacteriol. 181 (4), 1196-1202 (1999)
 MEDLINE 99138740
 PUBMED 9973346
 2 (bases 1 to 1587)
 REFERENCE Neuhaus, H.E.
 AUTHORS Direct Submission
 TITLE Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck, Biology / Plant Physiology, Barbarastr. 11, D-49069 Osnabrueck, GERMANY
 FEATURES
 source location/Qualifiers
 1..1587
 /organism="Chlamydia trachomatis"
 /mol_type="genomic DNA"
 /db_xref="taxon:813"
 1..1587
 /gene="np1"
 1..1587
 /gene="np1"
 /function="ATP/ADP transport protein"
 /codon_start=1
 /evidence=experimental
 /transl_table=1
 /product="nucleoside triphosphate transport protein 1"
 /protein_id="CA39534.1"
 /db_xref="GI:4538860"
 /db_xref="GGA:0956V3"
 /db_xref="SPTREMBL:O956V3"
 /translation="MTQNAKPEPKGRSEPMPTIHEELKVLDPMLPFISNYTTLVDFKDTLIVAPSGGAELPIKLMLVBPVAVFMILIYALSLINKQALFVAVLSFVDFPAPVIVYPCRHILHPTAFADTQSLIPGFMGFIAMLRNMTFAVYVISELWGVLSLMFMGFAIEIRISAKRFVAFVGVAVALLISGPALIMSKLASISGEVDPMVSLYPLMAMFLGCAITIAACVMNNRYVLDPDPVNPDELKAKSKDKMSGSEFVSLRSPNYMLIALIVICYGICINLYVTMKSQKXQDPNDVSAFMGNSFMTGVSVFVMLFISGNVIRFRGMLTGLAVTITMVLVTAIVFAVFIYFDHATGLVAACTPLMAVVGAIONITISKTYALFDATKELAVITPDDOKVKGKALIVAAVFKSGGSLIOGGLVVGSGSIGAMTPFLAVLAFIIMWLTSAITKLNKLFLAASAKQEDLAEEAAAEKASSAAKESAPALEIGVS"

Neleam, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S., Eisen, J. and Fraser, C.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On or before Jun 1, 2000 this sequence version replaced gi:7190382,
gi:7190372.

FEATURES
source location/Qualifiers
1..23773
/organism="Chlamydia muridarum"
/mol_type="genomic DNA"
/db_xref="taxon:83560"
106..1842
/locus_tag="TC0330"
106..1842
/locus_tag="TC0330"
/note="identified by match to PFAM protein family HMM
PF03000"
/codon_start=1
/transl_table=11
/product="protein export protein, FHIPF family"
/protein_id="AAF73547.1"
/db_xref="GI:8163204"
/translacion="MNHRGSSWKIVTIPICILFTLVLPRLMIDFGVCANLACSLS
IFWVSLRSSASARIIPSPSLLYCLILRLGLNASTRWLASGMASPLIPALGFEFL
GSLPALTVCILLPFINPLVITKMERIAEVRARARSESILPGKMSLADISAGRIGS
SRASYKNSLSESPYFSAMGEVRFVYGDAMSIVLLGVNSLALPTFGRYGYNNM
LFLVDALVSOVPALITSCSAITLAKYGEESLQHLVEYTEOCROSLFITALILCG
MAFIPGARKALILGFSVLLLGKSPYSESLFQKERIELLPKRGKENVVSYRAA
RHQIYOELIEFPEOGSLIEETIPLKISNSIVPLRALVEQVSIHLFKLADIL
FGSOAKSLSEIVSVRYNLGKQIGLSMNRDVEVITVDSLVEQFVDSQAKFVLD
NEKVSVQKDLIRBEGENFRAIVTCSERKELKRVDPFPDVLVAHSELPEEMPT
LIGAVSDVSLG"
1950..2711
/locus_tag="TC0331"
1950..2711
/locus_tag="TC0331"
/note="similar to GB:M57464, GB:X15262, GB:X12949,
GB:D00617, SP:P07949, PID:337356, and PID:38275;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor, sigma-70 family"
/protein_id="AAF3194.1"
/db_xref="GI:7190373"
/translacion="MKTQDLADTWQLVYSTKEIHHRDFLEAYLPLVKNVAHRLASGM
PSHVYKIEDIVASGVGLVAVRFPDESKRFEYSALFIIRAAIIDLRKODWPRSY
YQANRLADAMDALRQTKPEPTDGLCEYNIHQEELSHWSSRPALVSLINDPS
CODEDEGLAEERLADERAENGQVIRKKEAISITTEALSLDEKERQVMAVYYDDL
VLKETGKLGISESRVSGIHSKALLKRLGTLSSLL"
2869..4107
/gene="lysS"
/locus_tag="TC0332"
2869..4107
/locus_tag="TC0332"
/note="similar to GB:X53587, SP:P16144, PID:2270924,
PID:2293523, and PID:33951, identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="tyrosyl-tRNA synthetase"
/protein_id="AAF3195.1"
/db_xref="GI:7190374"
/translacion="MOQLDLSLQKRLINDNSAGLESITAFVSAVYLGFDPTAPSLIHG
HMGICFLRLSAGITPILAVGATGIDGPKSRYERSLDOEVLNDSKIEVLA
ANVLDIRIVNADMGLSISWVDLROIGKFRIGSMARXAVKORVYSESGISYTF
SYLLQSDYDPAHLFKHSGVYVCCGSGSDMGNTISGIDYIRKKGAGLFTPLTTS
KGKRTGESGVWLDPELISPYELFQFLRLSQDELTKARMLTLDDDLVADGR
LENDPOAVKRYVAVIVKDVHAGLGAALATLTPFNKNGKNTSESLALVSGGS
INVARADVIGRMVDVVVLGFCSSKGBARLLIQRGLYVQDEPLIDEGSLDGTLYC
FDRYILLGQKKKKVLDLN"

4117..5556
/gene="pgd"
/locus_tag="TC0333"
4117..5556
/gene="pgd"
/locus_tag="TC0333"
/note="similar to SP:P29716, SP:P37074, and GB:X63355;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="6-phosphogluconate dehydrogenase,
decarboxylating"
/protein_id="AAF3196.1"
/db_xref="GI:7190375"
/translacion="MAPADIGLGLAMGNULVNMTHGPAVSVNRSPEKTEPIK
EHEGNISLUGFTAIIEFVQSLKPRKIMINKAGAPDENTISLPLEBGDILIDG
NSYLLDERRYIDLKKGGLIFVGVSGEGRKQSGIMPGENIEMVYAPFQSI
AAQVDCQPCCSWIGTGAGHFVAVNHGIEYGIOLICETYEILKSRDLDSLQGINI
PFENWOTDLSNYLMAGAAVLTADENGVAVSTIIDLACQKGTGRVADAIKAGVP
MSLITESVILARYLSAMKEVRROARPEPVASLIYQPSASVILIDREARLAKIIS
YAGFNLKQISERBNVDLNGELALINRGCTIQSFLEKTIHQGFSCPDANSMIQ
DYFNKVLNSETGRRAILHAGAGVAILPCLASALAFYDGIKTENSLFLVQGLDYF
GAGYERQDRPREPFYHTDMLGSKNASRM"
complement (5624..7432)
/gene="lepA"
/locus_tag="TC0334"
complement (5624..7432)
/gene="lepA"
/locus_tag="TC0334"
/note="similar to GB:L22005, SP:E49427, and PID:388309;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein LepA"
/protein_id="AAF3197.1"
/db_xref="GI:7190376"
/translacion="MKPKYIENIRNFISIIAHIDHGSTIADRLLESTITEOREMEQ
LIDMDLERGRTIKAHPTMTVEYEGEYVLTIDTPHVPSEVSSSLACBGA
LIVDAQGOAOGLSANVVALRDLLETIPVUNKIDPAOPAKIKOIEFIDGS
NATACSAKRTQGIPELIESITIRVPPPKSETEBELKALITDSHYDPVGMVYRVIS
GEIKKGRTITENSTKSSFEVLIGAVLEPATLIESLRAGQVGFPLANKKXVDVI
GATVTVYKHPAKESPLBGFKEIKPVFAGIYPIDSDPDLKDALGRLOJLNDALTEQ
ESSHSLGFGRCFLGLHLIELFEERSREFDIIATAPSVYKVLKNGKTLFDIN
PTAVPDPALEHNEEPMVHNIIITPOEYLSNIMSLCDKXGICLKTMDLOHVLVSY
ELPLNETVSDPNDKSKSVTKGSPYRGDYRGALIKELIINDETUNDAFSCIVHR
DKAESKRSICEKLVADITPQLKPIQALINKKTIARETIRLAKVTKCYGGDIT
RKRLWEKQKKRMRKEFGKVSIPTAFAVEVLMKE"
complement (7617..9206)
/gene="t1c-1"
/locus_tag="TC0335"
complement (7617..9206)
/gene="t1c-1"
/locus_tag="TC0335"
/note="similar to GB:D13642, and PID:285999; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ADP, ATP carrier protein"
/protein_id="AAF3198.1"
/db_xref="GI:7190377"
/translacion="MTQAEKPFKLSFLPMIHMBLKKVLPMLMFCISFNYTIL
RDKDLIVAPSGSAAIPFLKMLVVSAAVFMILYANKLSIKOALFYAVLSPF
VDFPALPLVIVPYRHLIHPTRADLOLITLPGFPLAMENMPPAAFYVLSLWG
SVLSTLMFQFANETIKISBAKRFYALFGVANVALLISPAIVMSKLRASIGECVD
PWGVTVYFLMAMFLCSAIIAACYYMMNRVLDPRFYNDELKAKSKPKXMGESF
SYLRSPYMLLALLVICYGVCLNLEVSQKQKQFPNPNDSAFMGVFSFWTGV
SVVMVLEGNITIRREGMLTGAIVTVVMVLTGAIFFALVIFPDHATGVALAGTTP
MLAVVAGVONITISKSTKVALPDATKEMAYIPLDQSKVGAIDVAVARPKSGGS
LIDQGLLVCGSISANTPLAVVLPALINWLTSAIKKLKFLAASAAEQELAEATA
AAEKESPAKESVPAIBGVS"
complement (9553..10029)
/locus_tag="TC0336"
complement (9553..10029)

```

/locus_tag="TC0336"
/notes="similar to GB:M63957, SP:P30720, PID:151242, and
PID:99193; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA039199.1"
/db_xref="GI:7190378"
/transl_table="MATR01TIOEIR01ITKASXVGSKENDLCVLPQPGVWH
HFTLKAKHSAPBOLLMKLTFTVDSMTSPSINKRPAKRSKRKPVPATKIDIR
VLEIARQVGDLDIARSPKPLPSLRRLIRIRNIIVSTEINAVVLEADPTE"
/complement(10169..10270)
/locus_tag="TC0337"
/complement(10169..10270)
/locus_tag="TC0337"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA039200.1"
/db_xref="GI:7190379"
/transl_table="MOTNLENNKHFIPECPNOKISMRRLKXOKPL"
/locus_tag="TC0338"
/locus_tag="TC0338"
/locus_tag="TC0338"
/locus_tag="TC0338"
/notes="similar to GB:X63368, GB:X63368, SP:P25686,
PID:32469, and PID:32470; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1

```

Alignment Scores:

```

Pred. No.: 1.8e-179 Length: 23773
Score: 2143.00 Matches: 408
Percent Similarity: 89.04% Conservative: 55
Best Local Similarity: 78.46% Mismatches: 49
Query Match: 81.48% Indels: 8
DB: 1 Gaps: 3

```

US-09-869-433-2 (1-515) x AE002301 (1-23773)

```

QY 1 MetThrIysThrGluGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
DB 9206 ATGACTCAAAACCGCGGAAAACTTTGGAAAATTGGCCTCTTCCCTTGGCCCAATCAC 9147
QY 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
DB 9146 ATGCATGAGCTTAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGATTTTCAAT 9087
QY 41 TyrThrValLeuArgAspThrLysAspThrIleuIleValGlyAlaProGlySerGlyAla 60
DB 9086 TACACGATTTTGAAGATACAAAAGACACTCTTAATGTCACAGACAGCGCTCGAGCA 9027
QY 61 GluAlaIleProPheIleLysPheTrpLeuValAlaProCysAlaIleIlePheMetLeu 80
DB 9026 GAACCAATCCCTTATTAAGTTGGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8967
QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValAlaThr 100
DB 8966 ATCAACCAAAAGTTAAGCAATATTTTGAAGAAAGCGCTCTTTCTATGCAAGTCTCT 8907
QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
DB 8906 CCATTCCTGTGTTTTTGAAGCTGCTCCCTTGGTGAATTAACCATATCGCATATATCT 8847
QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
DB 8846 CACCAACAGACTTCGCTGATACACTACAGCCATTCCTCTTGGGTTTATGGCTTC 8787
QY 141 ValAlaIleLeuArgAspThrThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly 160
DB 8786 ATTCGATGCTACCAACTGGAATTTTGAAGCTGTTTATGTTCTTTCGAACTTTGGGA 8727

```

```

QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleGly 180
DB 8726 AGCTAATGCTTTTCGCTAATGTTCTGGGATTTTGCANATATACCTTAATATTAAGCAA 8667
QY 181 AlAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
DB 8666 GCTTAAGCTTTTCAAGCTCTTTTCGAGGTTGAGAGCTAAGCTAGCTCTTAATTTCCGGT 8607
QY 201 ArgAlaIleValIleTrpAlaSerLysLeuArgAlaSerAlaSerGluGlyValAspProTrp 220
DB 8606 CCAAGCTTTGTTGGCTCTTAATTAATTAAGCGCAAGCTTAAGTGAAGCGTGAAGCCCATGG 8547
QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
DB 8546 GGAATTACTCTTACTTCTTAATGCAATGCTTCTGTTCTTCTGTCATTAATAGCTGCC 8487
QY 241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGly 260
DB 8486 TGCTACGAGTGAATGCAATGCTTAAGCTTCTTAAGTCTTAAGTCTTAATTAAGCGAGAG 8427
QY 261 MetGlnLysGlyLysIleGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
DB 8426 CTA---AAAGCTAAAAA---CTAAACCTAAGATGAGCATGGCGAGAGCTTCTCTTAT 8373
QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
DB 8372 CTGTAAAGATCCCTTAACAGCTCTTCTTAAGCTCTGCTGCTAATTTGCTAAGGAGTTTGC 8313
QY 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn 320
DB 8312 ATCAACTTCTTAAGTAACTTGAAGTGAAGAGCCAGCTCAAAATGCAATTCCTTAATCCAAAT 8253
QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValAlaSerValLeuIle 340
DB 8252 GAATACAGCCCAATCAATGAGCAATCTCTTCTGCAAGGAGGTGATCCGATTCGTT 8193
QY 341 MetLeuPheValGlyGlyAsnValIleArgLysPheGlyTrpLeuThrGlyAlaLeuVal 360
DB 8192 ATGCTGTTCAATGGTGAATGATATTCGCTGTTGGCTGTTGAACAGCGCTTATGTT 8133
QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
DB 8132 ACACCTGTACATGATCCCTGTAACAGAGCATCTTCTTCTCTGTTATCTTCAAGAT 8073
QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValAla 400
DB 8072 CATGCAACAGATGTTGTTGAGCTTGGGAACAATCTTAAATGCTAAGCGGTGTTGTT 8013
QY 401 GlyAlaIleGlnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
DB 8012 GAGGCGCTCAAAATATTTCTTCCAAATCACTAAATATCTCTTGTATGCAACTAAA 7953
QY 421 GluMetAlaTyrIleProLeuAspGlnGluGlyLysValAlaLysGlyLysAlaIleAsp 440
DB 7952 GAGATGCTTACATCCATTCATGATCAAGACAAAGTTAAAGAAAGCCGCTATTTGAC 7893
QY 441 ValValAlaAlaArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuVal 460
DB 7892 GTTGTGCTCTGATTTGTTAATCTGGGGAGATCTTGAATCAACAGAGCTCTGATTT 7833
QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaAlaIleLeuPheIleIle 480
DB 7832 GTTTGGGAAGCATCACTGATGATCTCAATCTTACCGGTACTCTTTCGCAATTAAT 7773
QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
DB 7772 ATGTTTGGTTAATCAATCTGCAACCAACTTAATTAATTTGTTTCTTACGCTGCT 7713
QY 501 LysGlnGlnGluVal-----AlaGlnGluAspSerAlaProAlaSer 514
DB 7712 AAAGAACAAAGCTGGAGAGAGCCACAGCAGCAGCTGAGAAAGAGCTTCTCTGAGCGC 7653

```


KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-09-869-433-2 (1-515) x AX660386 (1-1851)

Oryza sativa
Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

1
Glazebrook, J., Briggs, S., Cooper, B., Goff, S.A., Moughamer, T.,
Katsagiri, P., Kreps, J., Provart, N., Rhee, D. and Zhu, F.
Plant disease resistance genes
Patent: WO 0300906-A 743 03-JAN-2003;
Syngenta Participations AG (CH)
Location/Qualifiers
1. 1851
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"

401 a 453 c 509 g 487 t 1 others

4.81e-113 Length: 1851
1379.50 Matches: 275
71.71% Conservative: 90
54.03% Miscellaneous: 135
52.45% Indels: 9
Gaps: 6

7 LysProhegilyLysLeuArgSerPheLeuThrProIleHisThrHisGluLeuLys 26
235 CAGCGGAGGGGAGG-----AAGTTCTCT--GCGGTGAGGTGAAGACCTGAAGAAG 285
27 ValLeuProwetPheLeuMetPhePheCysIleThrPheAsnYrThrValLeuArgAsp 46
286 ATCGTCGCGCTCGGGTGTGATGTTCTTCTGCACTTTGTTCAACATCAAGATCTCGGAGC 345
47 ThrLysAspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIleProPheIle 66
436 ACCAAGACGTCGTCGTGTACCGCCAGAGGGGAGCGCGGAGATCATCTCCGTTCTTC 405
67 LysPheThrLeuValValProCysAlaIleIlePheMetLeuIleYrAlaLysLeuSer 86
406 AAGACGTGGGCAACCTGCCATGCGCATCGGCTTCACTCTCCACCAACCACTTCC 465
87 AsnIleLeuSerIlyGlnAlaLeuPheYrAlaValGlyThrProPheLeuIlePhePhe 106
466 AATGTGCTCTCCAGAGGGGCGCTCTTCAACCGTCATCTCCCTTCATCGCTTCTTC 525
107 AlaLeuPheProThrValIleYrProLeuArgAspValLeuHisProThrGluPheAla 126
526 GGCCTCTTGGCTTCGTGCTCTTACCTCTCAGAAATGATGATCAACCCACCGCGCTCGCC 585
127 AspArgGlnGlnAlaIleLeuProProGlyLeuLeuGlyLeuValAlaIleLeuArgAsn 146
586 GACAACTCTCTGCGGCGCTCGCCGAGCTTCTTGCGCCCGCGCATTTCTAGAGATT 645
147 ThrThrPheAlaAlaPheYrValLeuAlaGluLeuThrGlySerValMetLeuSerLeu 166
646 TGGAGCTTTGCTTCTTCTATGTCATGCGGACGATGAGGGGCAACGTCATCTCGGTC 705
167 MetPheThrGlyPheAlaAsnGluIleThrLysIleHisGluAlaIlePhePheYrAla 186
706 CTCCTTCGGGCTTCGCTAACCAAGATTACTACGTTGAGGAAGAAAGAAATTTACCT 765
187 LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValThrAla 206
766 CTATTTGCTCTTGGGCTTAATATGCGCTCACTCTTTCTTGCGCGCTACTGTAAGATTTC 825
207 SerLysLeuArgAlaSerValSerGluGlyValAspProThrGlyIleSerLeuArgLeu 226
826 TCAAATCTGGCCAGACATTTGGTCCAGGATATTGACGATGGAGGATGTTCTTTAAAGGA 885

Ory 227 LeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerYrTrpTrpIleAsn 246
Db 886 ATATGAGCTTAGTGGGCTTCTTGGACTGTGATCATCTTCCATCTATTTGGGGGTAAC 945
Ory 247 LysAsnValLeuThrAspProArgPheYrAspProGluGluMetGlnLysGlyLys 266
Db 946 AAGTTGTTTGAATATATCTTCTCTT-----CCGAACTGACCGTAAAGAAAG 999
Ory 267 GlyAlaLysProLysMetAlaMetLysAspSerPheLeuYrLeuAspArgSerProYr 286
Db 1000 GAT---AAACCTAAGCTTGAAGAAAGAGACTTAAAGTCTCTCTCTCAAGATAT 1056
Ory 287 IleLeuLeuLeuThrLeuLeuValIleAlaYrGlyIleCysIleAsnLeuIleGluVal 306
Db 1057 GTAAGGACCTTCTCTACTTATGAGTGGCAATGAGCATCGATCAACCTTGTGAGCTC 1116
Ory 307 ThrTrpLysSerGlnLeuLysLeuGlnYrProAsnMetAsnAspYrSerGluPheMet 326
Db 1117 ACATGAAATCGAAGCTCAAGGCAAGCTTCCCTAGTCCAAATAGATTTCACTTTCATG 1176
Ory 327 GlyAsnPheSerPheThrGlyValValSerValLeuIleMetLeuPheValGly 346
Db 1177 GGTGATTTCTCAACCGCTACTGCAATGCTTCACTTCAATGATGCTG--TTAGCCCGG 1233
Ory 347 AsnValIleArgLysPheGlyYrThrLeuThrGlyAlaLeuValThrProValMetValLeu 366
Db 1234 ATCATCTCCGAAATTCGCTGGAGAGTGGCATGATGATCAACCCACAGTTTGTG 1293
Ory 367 LeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal 386
Db 1294 CTGACGTGCGCTCGGTTCTTCTCTCCGATTTCTATTGGGAGCGCTCACTCCATGCTT 1353
Ory 387 AlMetPheGlyThrThrProLeuMetLeuValValValGlyAlaIleGlnAsnIle 406
Db 1354 GCCAATATGGGCAATGACCCCTCTTCTGACAGATTTATGATGGTGACATTCAAGACATA 1413
Ory 407 LeuSerLysSerThrYrAlaLeuPheAspSerThrYrGluMetAlaYrIlePro 426
Db 1414 TTGAGCAAGGTGCAAAAGTATGTTTATGATTCATGCAAAAGAAAGGCTTCATTTCT 1473
Ory 427 LeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAspValValAlaAlaArgPhe 446
Db 1474 TTGATGAAATATGAAAGTGAAGGGAAGGAGCGCATTTATGTTGCTTCCAAACCTTTC 1533
Ory 447 GlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySerIleGly 466
Db 1534 GGGAACTGCTGGAGCTTATTTCAACAGTTCATGATCTTCACTTGGCTTTAGCT 1593
Ory 467 AlMetThrProYrLeuAlaValIleLeuLeuPheIleIleAlaIleThrPheValSer 486
Db 1594 AACTCCACCCCTTACCTTGTGTCATACGTTGATGTTGTTGATGATGCTTCAAGCGCG 1653
Ory 487 AlaThrLysLeuAsnLysLeuPhe-----LeuAlaGlnSerAlaLeuLysGluGln 504
Db 1654 GCAAGTCTACTAAGCAACGAATTTCTAGCTTGGCTTGAAGAAATCTCAAAAGGATATG 1713
Ory 505 ValAlaGlnLysSerAlaProAla 513
Db 1714 TCTGCAAAAGAGAAAGTATGATCTTCC 1740

RESULT 12
AX654737 1872 bp DNA linear PAT 22-MAR-2003
LOCUS
DEFINITION
Sequence 4607 from Patent WO0300898.
AX654737
ACCESSION
VERSION
AX654737.1 GI:29157551
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
 Karaganti, F., Qian, S., Tao, X., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
 TITLE Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 0300898-A 4607 03-JAN-2003;
 Syngenta Participations AG (CH)
 FEATURES
 SOURCE
 1. 1872
 /organism="Oryza sativa"
 /mol_type="genomic DNA"
 /db_xref="taxon:4530"
 BASE COUNT 384 a 545 c 497 g 442 t 4 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.14e-109 Length: 1872
 Score: 1338.50 Matches: 266
 Percent Similarity: 70.38% Conservative: 88
 Best Local Similarity: 52.88% Mismatches: 141
 Query Match: 50.89% Indels: 9
 Gaps: 3
 DB: 6
 US-09-869-433-2 (1-515) x AK654737 (1-1872)

Db 958 -----TCAGTCGGCCACGTCGTGACGGGAAAGAAAGATAGCCAAACCTG 1005
 Oy 273 AsmMetLysAspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuThrLeu 292
 Db 1006 AGTATGGAGAGAGTTTGAAGAGTCTGTGTGATCTCGTATGTGAGGAGATCTTGCCACC 1065
 Oy 293 LeuValIleLeuArgTyrGlyIleCysIleAsnLeuIleGlyValThrTyrPheSerGlnLeu 312
 Db 1066 CTGGTCGTGCTTATGTGATATGACATTAACCTTGTGAGAGGTGACATGAGATGAAAGCTG 1125
 Oy 313 LysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTyr 332
 Db 1126 AAGGACAAATACCCAAACCCCTAATGATATTTCTTATCATGGGGAGATTTCTGCACGCT 1185
 Oy 333 ThrGlyValValSerValIleMetLeuPheValGlyValAsnValIleGlyPhe 352
 Db 1186 ACTGGCATACCAATTTACATGATGTG---TTAGGGAGAGTAAATCTTGAAGAGTTT 1242
 Oy 353 GlyTyrPheThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhe 372
 Db 1243 GGTGGGAGTCCGACGCTTACATACCCCTGCAGTGTGCTTCTCATGTGAGTTGATTC 1302
 Oy 373 PheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThr 392
 Db 1303 TTCTCATGATTTTATTCGTTGAGCCATGATGCTCTTATGAGCCACATTTGGAATGACA 1362
 Oy 393 ProLeuMetLeuAlaValIleValGlyAlaIleGlnAsnIleLeuSerIleSerThrLys 412
 Db 1363 CTTTGGCTCTGCGGTCTATGTTGGGGCAATGAGAACATTTTCAGTAAGAGTCCGAG 1422
 Oy 413 TyrAlaLeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGlnLys 432
 Db 1423 TACAGTTGTTGACCTTGTATGAAATGCAATCTCTTGTGCGAGACATGATAG 1482
 Oy 433 ValIleGlyLysAlaAlaIleAspValIleAlaAlaAlaArgPheGlyLysSerGlyAla 452
 Db 1483 GTCAAGATTAAGAGAGCAATTGATGTGTGTGCAACCATTTGGGAAAGTGTGAGGGCC 1542
 Oy 453 LeuIleGlnGlnLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeu 472
 Db 1543 TTGATCCAGAGTTCATGATCTTATGATGATCTCTCCAAATTTACACCTTACTG 1602
 Oy 473 AlaValIleLeuPheIleIleAlaIleTyrLeuValSerAlaThrLysLeuAsnLys 492
 Db 1603 GAGAGATTTCTATTTGATTTGTTCTTGCTGCTGGTGTCTGAAGTCCCTTGACTCA 1662
 Oy 493 LeuPheLeuAlaGlnSerAlaLeuLysGluGlnGluValAlaGlnGluAspSerAlaPro 512
 Db 1663 CAGTTC-----TCCCACTGGCAAGCAAGAGCTTGAAAGAAATGCTGAAA 1713
 Oy 513 AlaSerSer 515
 Db 1714 GCAAGAGCG 1722
 RESULT 13
 AX505999 1875 bp DNA linear PAT 27-SEP-2002
 LOCUS Sequence 694 from Patent WO0216655.
 DEFINITION AX505999
 VERSION AX505999.1 GI:23387236
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Arabidopsi thaliana
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1
 AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
 TITLE Stress-regulated genes of plants, transgenic plants containing
 same, and methods of use
 JOURNAL Patent: WO 0216655-A 694 28-FEB-2002;
 The Scripps Research Institute (US); Syngenta Participations AG

[illegible]

JOURNAL
Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory
(SIGMA). Plant Biology Laboratory, The Salk Institute for

Direct Submission
Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory
(SIGAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037
USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RFL CDNA (RAFL CDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Saitou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shim, P., Ban, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

source
Location/Qualifiers
1. 2110
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAFL09-51-C18(R14701)"
/note="ecotype: Columbia"
1. 55
56. 1930
/note="adenine nucleotide translocase"
/codon_start=1
/product="At1G80300/F516_5"
/protein_id="F516246.1"
/db_xref="GI:16226734"

5'UTR
CDS

3'UTR
BASE COUNT 525 a 454 c 502 g 629 t
ORIGIN
1931..2110
3'
ATGTTGGTATGCTCTCTACCTAACTCCAAATGTTCTCCAAAGAGCTGTTT
608 TACACTGTTATGTTCCCTTTCATCATCTACTTGGGGCTTGGTTGTCATGTAACCT
116 LeuArgValLeuHisProThrGluPheAlaAspArgLeuAlaIleLeuPro 135
668 CTAGCAACATATATTCACCCGGAAGCTCTGCAGATAGCTCTTCAACCTCGGCCA 727
136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnThrPheAlaAlaPheValLeu 155
728 AGATTACATGAGTCTCTATGCAATATTCGAGATTGAGATTCTGTTTATGTTATG 787
156 AlGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGlu 175
788 GCTAGGTTTGGGGATGATGGTGGTGGTCTCTCTCTGGGGCTTGGTAAACAGATC 847
176 ThrLeuIleHisGluAlaValArgPheValAlaLeuPheGlyIleGlyAlaAsnIleSer 195
848 ACAACTGTGATAGCAAGCAAAATCTATCTTGTTCGACTGGACCAATGTTGCA 907
196 LeuLeuAlaSerGlyValArgAlaIleValAlaSerLeuArgAlaSerValSerGlu 215
908 CTGATTTTCTCAGGAAGAACCGTAATACTTCTTACTTGAAGAAATCTTGCTCT 967
216 GlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
968 GGAATTGACCGCTGGGAGATGTTGTTGTAAGCAATGATGATGATGATGATGATG 1027
236 LeuValLeuMetAlaSerValTrpTrpIleAsnValAsnValLeuThrAspProArgPhe 255
1028 CTGCGCATTTGTTCTCTCTATGATGGTGAATGATATTTCTCTTCCAAACCGTATG 1087
256 TyrAspProGluGluMetGlnGlyValGlyValAlaValAspProValMetAsnMetIys 275
1088 AAGAACAGAAAGAG-----AAGCAAGATGGGAAGCATG 1123
276 AspSerPheLeuValTrpLeuAspArgSerProValIleLeuLeuValIle 295
1124 GAAAGCTTGAAGTCTTGGTATCATCATCACTAATAGATCTTGTACTTATGCTGTG 1183
296 AlaTrpGlyIleCysIleLeuLeuIleGluValThrTrpValSerGlnLeuValLeuGln 315
1184 GCATAGGATTTATGATCAATCTTGTGAAAGTCAATGAAATCAAGCTTAAGCTCAG 1243
316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrpGlyVal 335
1244 TTCCCTAGCCCGAATGATGATCTCAGCATTTATGGAGACTTCTCAACCTGACCGGTGT 1303
336 ValSerValLeuIleMetLeuPheValGlyAsnValIleArgIysPheGlyTrpLeu 355
1304 GCAACATTTACATATGATCTT---CTCAGCCATTAAGTATTAAGTATGTTGGGA 1360
356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
1361 GTAGCTGCAAAAGATCAACCCCACTGTTCTGCTATGATGATGCTGCTTCTCTCTA 1420
376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
1421 ATATGTTGGCGGCCCATTCGACACCATGTTGGCAAGCTTGGTATGACACCGTACTT 1480
396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerIysSerThrIysValAlaLeu 415
1481 GCAGCTGTATGTCGTCCTTCAAGATATCTTCAAGCAAGATGTCAGAGTCAAGCTTG 1540
416 PheAspSerThrArgGluMetAlaValIleProLeuAspGlnGluIysValValGly 435
1541 TTGACCTCTTCAAAAGAAATGCTTATATCCATGATGATGAGCAACAAAGTTAAAGGC 1600
436 LysAlaAlaIleAspValValAlaAlaArgPheGlyValSerGlyValAlaLeuIleGln 455
1601 AAGCTGCGATTAACGTCGTCTCAACCATTAAGGAAATCAAGGGAGCTTTATATCAG 1660

Alignment Scores:

Pred. No.: 7.67e-109 Length: 2110
Score: 1333.00 Matches: 259
Percent Similarity: 69.60% Conservative: 89
Best Local Similarity: 51.80% Mismatches: 140
Query Match: 50.68% Indels: 12
Gaps: 3

US-09-869-433-2 (1-515) x AF428316 (1-2110)

QY 16 LeuTrpProIleHisThrHisGluLeuValValLeuProMetPheLeuMetPhePhe 35
Db 368 ATTTTCGCTGTGAGGTTGCACTTGAATAAGATATTCCTTTAGGATTTGATGTTCTTT 427
QY 36 CysIleThrPheAsnValTrpValLeuArgAspThrIysAspThrLeuIleValGlyAla 55
Db 428 TGTATTTCTTTCAATTCACAAATCTGAGGATTCACAAAGATGCTTGTGTGTCGCGC 487
QY 56 ProGlySerGlyAlaGluAlaIleProPheIleValSerPheTrpLeuValAlaProCysAla 75
Db 488 AAAAGAAATTTCTGCTGATTTATACCTTTCTTGAAGCTTGGGGAATCTTCCATGCGC 547
QY 76 IleIlePheMetLeuIleValAlaValLeuSerAsnIleLeuSerIysGlnAlaLeuPhe 95

Db 548 ATGCGTTATGCTCTCTACCTAACTCCAAATGTTCTCCAAAGAGCTGTTT 607
QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
Db 608 TACACTGTTATGTTCCCTTTCATCATCTACTTGGGGCTTGGTTGTCATGTAACCT 667
QY 116 LeuArgValLeuHisProThrGluPheAlaAspArgLeuAlaIleLeuPro 135
Db 668 CTAGCAACATATATTCACCCGGAAGCTCTGCAGATAGCTCTTCAACCTCGGCCA 727
QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnThrPheAlaAlaPheValLeu 155
Db 728 AGATTACATGAGTCTCTATGCAATATTCGAGATTGAGATTCTGTTTATGTTATG 787
QY 156 AlGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGlu 175
Db 788 GCTAGGTTTGGGGATGATGGTGGTGGTCTCTCTCTGGGGCTTGGTAAACAGATC 847
QY 176 ThrLeuIleHisGluAlaValArgPheValAlaLeuPheGlyIleGlyAlaAsnIleSer 195
Db 848 ACAACTGTGATAGCAAGCAAAATCTATCTTGTTCGACTGGACCAATGTTGCA 907
QY 196 LeuLeuAlaSerGlyValArgAlaIleValAlaSerLeuArgAlaSerValSerGlu 215
Db 908 CTGATTTTCTCAGGAAGAACCGTAATACTTCTTACTTGAAGAAATCTTGCTCT 967
QY 216 GlyValAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
Db 968 GGAATTGACCGCTGGGAGATGTTGTTGTAAGCAATGATGATGATGATGATGATG 1027
QY 236 LeuValLeuMetAlaSerValTrpTrpIleAsnValAsnValLeuThrAspProArgPhe 255
Db 1028 CTGCGCATTTGTTCTCTCTATGATGGTGAATGATATTTCTCTTCCAAACCGTATG 1087
QY 256 TyrAspProGluGluMetGlnGlyValGlyValAlaValAspProValMetAsnMetIys 275
Db 1088 AAGAACAGAAAGAG-----AAGCAAGATGGGAAGCATG 1123
QY 276 AspSerPheLeuValTrpLeuAspArgSerProValIleLeuLeuValIle 295
Db 1124 GAAAGCTTGAAGTCTTGGTATCATCATCACTAATAGATCTTGTACTTATGCTGTG 1183
QY 296 AlaTrpGlyIleCysIleLeuLeuIleGluValThrTrpValSerGlnLeuValLeuGln 315
Db 1184 GCATAGGATTTATGATCAATCTTGTGAAAGTCAATGAAATCAAGCTTAAGCTCAG 1243
QY 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrpGlyVal 335
Db 1244 TTCCCTAGCCCGAATGATGATCTCAGCATTTATGGAGACTTCTCAACCTGACCGGTGT 1303
QY 336 ValSerValLeuIleMetLeuPheValGlyAsnValIleArgIysPheGlyTrpLeu 355
Db 1304 GCAACATTTACATATGATCTT---CTCAGCCATTAAGTATTAAGTATGTTGGGA 1360
QY 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
Db 1361 GTAGCTGCAAAAGATCAACCCCACTGTTCTGCTATGATGATGCTGCTTCTCTCTA 1420
QY 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
Db 1421 ATATGTTGGCGGCCCATTCGACACCATGTTGGCAAGCTTGGTATGACACCGTACTT 1480
QY 396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerIysSerThrIysValAlaLeu 415
Db 1481 GCAGCTGTATGTCGTCCTTCAAGATATCTTCAAGCAAGATGTCAGAGTCAAGCTTG 1540
QY 416 PheAspSerThrArgGluMetAlaValIleProLeuAspGlnGluIysValValGly 435
Db 1541 TTGACCTCTTCAAAAGAAATGCTTATATCCATGATGATGAGCAACAAAGTTAAAGGC 1600
QY 436 LysAlaAlaIleAspValValAlaAlaArgPheGlyValSerGlyValAlaLeuIleGln 455
Db 1601 AAGCTGCGATTAACGTCGTCTCAACCATTAAGGAAATCAAGGGAGCTTTATATCAG 1660

```

Qy      456 GlnGlyLeuLeuValIleCyseGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475
Db      1661 CAGTTCATGATCTTATTCCTTGGATCAGCAATTCAAGCCGCTATCTAGGAATGATC 1720
Qy      476 LeuLeuPheIleIleAlaIleTyrLeuValSerIleThrIleLeuValLeuPheLeu 495
Db      1721 TTGTTGGTTATTGTCAGTGGCGTGGTTAGTCAAGCTAAGTCGCTGAGGACAGTTTC--- 1777
Qy      496 AlaGlnSerAlaLeuIleGlnGluValAlaGlnGluIleAspSerAlaProIleAspSer 515
Db      1778 -----AACAGCTTGGCGGTCTGAAGAAGAGCTTGAGAGGAATGAGAGAGAGCTTCATCG 1831

```

Search completed: November 25, 2003, 12:09:54
 Job time : 6065 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:12:27 ; Search time 28 Seconds

(without alignments)
1768.817 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKTEKRPFKLRSLFMPH.....AQSLKQEVAAQEDSAPASS 515

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: PIR 76:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2617	99.5	515	2 E72089	ADP, ATP carrier p
2	2610	99.2	515	2 E86534	ADP/ATP translocas
3	2147	81.6	528	2 C71561	probable adp/atp t
4	2143	81.5	529	2 C81774	ADP, ATP carrier p
5	1333	50.7	624	2 E86834	adenine nucleotide
6	1330.5	50.6	618	2 G86288	probable adenine n
7	1306	49.7	631	2 T07420	ATP/ADP translocas
8	1220.5	46.4	589	1 S68205	ATP/ADP translocas
9	1076	40.9	498	1 A97710	ADP, ATP carrier pr
10	1056	40.2	498	1 J00026	ADP/ADP translocas
11	1025.5	39.0	540	2 C86567	ADP/ATP translocas
12	1025.5	39.0	540	2 B72056	ADP, ATP carrier p
13	976.5	37.1	501	2 B97750	ADP, ATP carrier pr
14	969.5	36.9	501	2 B71707	ADP, ATP carrier pr
15	956.5	36.4	540	2 B71503	probable adp/atp t
16	953	36.2	543	2 F81665	ADP, ATP carrier p
17	907.5	34.5	499	2 B97842	ADP, ATP carrier pr
18	907.5	34.5	511	2 B97783	ADP, ATP carrier pr
19	880.5	33.5	512	2 F71653	ADP, ATP carrier p
20	865	32.9	500	2 B71633	ADP, ATP carrier pr
21	827	31.4	507	2 B97765	ADP, ATP carrier pr
22	824	31.3	507	2 B71695	adp, ATP carrier p
23	229.5	8.7	441	2 H82642	hypothetical prote
24	173.5	6.6	928	2 F71541	hypothetical prote
25	161	6.1	918	2 B81695	conserved hypotet
26	158.5	6.0	925	2 A72096	ct334 hypothetical
27	157.5	6.0	925	2 B81573	conserved hypotet
28	157.5	6.0	925	2 E86527	CT334 hypothetical
29	151	5.7	1002	2 AF2363	hypothetical prote

30	145	5.5	540	2 B82219	transporter, BCCT
31	144	5.5	660	2 E81839	probable transmemb
32	144	5.5	675	2 E81101	transporter, BCCT
33	140	5.3	474	2 B83719	multidrug resistanc
34	137	5.2	536	2 T07946	cytochrome-c oxida
35	136.5	5.2	498	2 T14236	NADH2 dehydrogenas
36	135	5.1	441	2 C83615	probable MFS trans
37	134	5.1	975	2 E75212	oligosaccharyl tra
38	133	5.1	669	2 D64137	bet1 protein homol
39	132.5	5.0	463	1 GRECKK	nitrite extrusion
40	131.5	5.0	463	2 H90844	nitrite extrusion
41	131.5	5.0	463	2 G85702	nitrite extrusion
42	131.5	5.0	567	2 A99445	permease, multitu
43	131	5.0	434	2 AC2833	hypothetical prote
44	131	5.0	434	2 C97610	hypothetical prote
45	130.5	5.0	976	2 C71248	hypothetical prote
46	129.5	4.9	446	2 A10401	probable membrane
47	129.5	4.9	581	2 B90539	hypothetical prote
48	129	4.9	462	2 S11431	nitrite extrusion
49	129	4.9	484	2 T07782	cytochrome-c oxida
50	128.5	4.9	435	2 C89795	hypothetical prote
51	128.5	4.9	528	2 S26948	cytochrome-c oxida
52	128.5	4.9	665	2 H87468	ubiquinol oxidase
53	128	4.9	462	2 H90330	amino acid transp
54	127.5	4.8	505	2 B88206	protein F21D12.3 l
55	127	4.8	462	2 H90887	nitrite extrusion
56	127	4.8	462	2 A85720	nitrite extrusion
57	126.5	4.8	512	2 G69670	glycine betaine tr
58	126.5	4.8	526	2 S62763	cytochrome-c oxida
59	126	4.8	548	2 T51035	hypothetical prote
60	126	4.8	664	1 B32243	phosphotransferase
61	125.5	4.8	641	2 S44253	raffinose carrier
62	124	4.7	391	2 D71676	bicyclomycin resis
63	123.5	4.7	394	2 S39739	efflux protein hom
64	123.5	4.7	531	2 S78190	cytochrome-c oxida
65	122.5	4.7	465	2 AG0648	nitrite extrusion
66	122.5	4.7	512	2 E83060	iron (III)-transpo
67	122.5	4.7	1139	2 A10379	probable potassium
68	122	4.6	643	2 E69373	conserved hypotet
69	121.5	4.6	894	2 S45135	probable membrane
70	121	4.6	536	2 S63651	cytochrome-c oxida
71	120.5	4.6	472	2 F90078	hypothetical prote
72	120.5	4.6	1118	2 B83018	conserved hypotet
73	120	4.6	389	1 C64923	chloramphenicol re
74	120	4.6	787	2 H70374	NADH2 dehydrogenas
75	119.5	4.5	391	2 AG0533	probable drug effl
76	119	4.5	389	2 F90924	probable transport
77	119	4.5	389	2 B85773	probable transport
78	119	4.5	493	2 AG1773	drug-export protei
79	119	4.5	1131	2 A97650	hypothetical prote
80	119	4.5	1131	2 A82873	two component sens
81	118	4.5	462	2 AC0672	nitrite extrusion
82	117.5	4.5	504	2 E64118	sodium/proline sym
83	117.5	4.5	507	2 AB1707	glycine betaine tr
84	117.5	4.5	512	2 E89978	high affinity prol
85	117.5	4.5	744	2 AG1382	transport protein
86	117	4.4	493	2 AD1398	drug-export protei
87	117	4.4	684	2 E97943	Na+/H+ antiporter
88	116.5	4.4	434	2 H83327	hypothetical prote
89	116.5	4.4	522	2 S25956	cytochrome-c oxida
90	116.5	4.4	1126	2 A96032	probable two-compo
91	116	4.4	472	2 T34748	transmembrane tran
92	116	4.4	475	2 A70330	cytochrome c-type
93	116	4.4	635	1 A64162	conserved hypotet
94	116	4.4	684	2 D65308	hypothetical prote
95	115.5	4.4	397	2 D65066	hypothetical prote
96	115.5	4.4	1050	2 G70396	cation efflux syst
97	115	4.4	406	2 T43120	conserved hypotet
98	115	4.4	487	2 A40956	probable membrane
99	115	4.4	521	2 A32431	cytochrome-c oxida
100	115	4.4	541	2 A48337	cytochrome-c oxida

ALIGNMENTS

RESULT 1

ADP/ATP carrier protein CP0408 (imported) - Chlamydomonas pneumoniae (strain CML029 at E72089
 C/Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C/Accession: E72089; D81580
 R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999.
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A/Reference number: A72000; MUID:99206606; PMID:10192388
 A/Accession: E72089
 A/Molecule type: DNA
 A/Residues: 1-515 <ARN>
 A/Cross-references: GB:AE001619; GB:AE001363; NID:94376620; PIDN:AA018495.1; PID:9437662
 A/Experimental source: strain CML029
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: D81580
 A/Molecule type: DNA
 A/Residues: 1-515 <REA>
 A/Cross-references: GB:AE002202; GB:AE002161; NID:97189324; PIDN:AA938252.1; PID:9718933
 A/Experimental source: strain AR39, HL cells
 C/Genetics: A/Gene: act 1; CP0408
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.5%; Score 2617; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 8.7e-193;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MTKEEKPFGLRSLPFIHTEHFKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPGSGA 60
 1 MTKKEKPFGLRSLPFIHTEHFKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPGSGA 60
 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFALPPTIYPLRDVL 120
 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFALPPTIYPLRDVL 120
 121 HPTFADRLQAILPPGLGLVAILRNMTFAAFYVLAELMGSVMSLMWGFANETIKHE 180
 121 HPTFADRLQAILPPGLGLVAILRNMTFAAFYVLAELMGSVMSLMWGFANETIKHE 180
 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLNMAITVSGVLMA 240
 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLNMAITVSGVLMA 240
 241 SYMWINKNVLTDPFYNPEEMQKKGAKPKNNKDSFLYLRSPYILLTLVIAVIGIC 300
 241 SYMWINKNVLTDPFYNPEEMQKKGAKPKNNKDSFLYLRSPYILLTLVIAVIGIC 300
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 361 TPVAVLLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSK 420
 361 TPVAVLLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSK 420
 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGGALIQQGLVTCGIGAMTPYLAIVLPII 480
 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGGALIQQGLVTCGIGAMTPYLAIVLPII 480
 481 AIWVSATKLNKFLAQSALKQEVVAOEDSAPASS 515
 481 AIWVSATKLNKFLAQSALKQEVVAOEDSAPASS 515

RESULT 2

ADP/ATP translocase (imported) - Chlamydomonas pneumoniae (strain J138) E86534
 C/Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C/Accession: E86534
 R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishii, N.; Nucleic Acids Res. 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
 A/Reference number: A86491; MUID:2030349; PMID:10871362
 A/Accession: E86534
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-515 <STO>
 A/Cross-references: GB:BA000008; NID:98978723; PIDN:BA98559.1; GSPDB:GN00142
 A/Experimental source: strain J138
 C/Genetics: A/Gene: act 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.2%; Score 2610; DB 2; Length 515;
 Best Local Similarity 99.4%; Pred. No. 3e-192;
 Matches 512; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 MTKEEKPFGLRSLPFIHTEHFKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPGSGA 60
 1 MTKKEKPFGLRSLPFIHTEHFKVLPMFLMFCITFNTYTVLRDTKDTLIVGAPGSGA 60
 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFALPPTIYPLRDVL 120
 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFALPPTIYPLRDVL 120
 121 HPTFADRLQAILPPGLGLVAILRNMTFAAFYVLAELMGSVMSLMWGFANETIKHE 180
 121 HPTFADRLQAILPPGLGLVAILRNMTFAAFYVLAELMGSVMSLMWGFANETIKHE 180
 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLNMAITVSGVLMA 240
 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISRLNMAITVSGVLMA 240
 241 SYMWINKNVLTDPFYNPEEMQKKGAKPKNNKDSFLYLRSPYILLTLVIAVIGIC 300
 241 SYMWINKNVLTDPFYNPEEMQKKGAKPKNNKDSFLYLRSPYILLTLVIAVIGIC 300
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGWLTGALV 360
 361 TPVAVLLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSK 420
 361 TPVAVLLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSK 420
 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGGALIQQGLVTCGIGAMTPYLAIVLPII 480
 421 EMAYIPLDQEQKVKGAIDVVAARFGKSGGALIQQGLVTCGIGAMTPYLAIVLPII 480
 481 AIWVSATKLNKFLAQSALKQEVVAOEDSAPASS 515
 481 AIWVSATKLNKFLAQSALKQEVVAOEDSAPASS 515

RESULT 3

Probable adp/act translocase - Chlamydia trachomatis (serotype D, strain WJ3/Cx) C71561
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C/Accession: C71561
 R/Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia traci
 A/Reference number: A71570; MUID:9900809; PMID:9784136
 A/Accession: C71561
 A/Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-528 <ARN>
 A:Cross-references: GB:AE001281; GB:AE001273; NID:g3328454; PIDN:AA067656.1; PID:g332845
 A:Experimental source: serotype D, strain UM-3/CX
 C:Genetics:
 A:Gene: CT065
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 81.6%; Score 2147; DB 2; Length 528;
 Best Local Similarity 78.1%; Pred. No. 8.8e-157;
 Matches 410; Conservative 52; Mismatches 49; Indels 14; Gaps 3;

QY 1 MTKTEKPFGLKRSFLMPIHTEHLKYLPMFLMFCITFNVTYLRDRTDITLIVAGPSGA 60
 DB 1 MTOAEKPFGLKRSFLMPIHTEHLKYLPMFLMFCISFNITLRDRTDITLIVAPSSGA 60
 QY 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPLRDVL 120
 DB 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPCRHIL 120
 QY 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAAFYVLALMGSVMSLMFPGFANEITKISE 180
 DB 121 HPTAFADTLQAIIPSGFMGLFAMLRNMTFAAFYVLSELMSGLMFMGFAEITKISE 180
 QY 181 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPMGISRLIMAMTIVSGVLMA 240
 DB 181 AKRFYALFGIGANVALISGRPAITWSKLRASLSEGVDPMGISLYFLMAMFLCSCAITAA 240
 QY 241 SYWINKNVLTDPRFYNPEEMQKKGAKPKMNKDSFLYLDSPYILLTLIVAYGIC 300
 DB 241 CYWMNRVYVLTDRFYNPAEL-KAKK-SKPKMGSEFSYLLRSPLYMLLALIVICGYC 298
 QY 301 INLEVTWKSQQLQYRNNDYSEFGNFSFTGVSVYLMFLVGNVIRKFGMLTGALV 360
 DB 301 INLEVTWKSQQLKQKQFNPNDYSAFMGNFSFTGVSVFVWMLFGVNVIRKFGMLTGALV 358
 QY 299 INLEVTWKSQQLKQKQFNPNDYSAFMGNFSFTGVSVFVWMLFGVNVIRKFGMLTGALV 358
 DB 361 TPWVVLITGVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONISKSTKVALFPSTK 420
 QY 359 TPIMVLTGAVFPALVIFRDHATGLVVALGTTPLMLAVVGAIONISKSTKVALFPDARK 418
 DB 421 EMAYIPLDQKQKGAIDVVAARFGSGGALIOGSLIVCGSIGAMTPYLAIVLFI 480
 QY 419 EMAYIPLDQKQKGAIDVVAARFGSGGSLIOGSLIVCGSIGAMTPYLAIVLPAII 478
 QY 481 AIMVLSATKINKLFLAQSALKEQEV-----EDSAPA 513
 DB 479 MWVLTSATKINKLFLAASAKKEQELAEATAAKESASPAA 523

RESULT 4

ADP: ATP carrier protein TC035 [imported] - Chlamydia muridarum (strain Nigg)
 C/Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C/Accession: C81714
 R/Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, W.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A/Accession: C81714
 A/Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-529 <TGT>
 A:Cross-references: GB:AE002301; GB:AE002160; NID:g7190372; PIDN:AAF39198.1; PID:g719037
 A:Experimental source: strain Nigg (Mopn)
 C:Genetics:
 A:Gene: TC035
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 81.5%; Score 2143; DB 2; Length 529;
 Best Local Similarity 78.5%; Pred. No. 1.8e-156;
 Matches 408; Conservative 55; Mismatches 49; Indels 8; Gaps 3;

QY 1 MTKTEKPFGLKRSFLMPIHTEHLKYLPMFLMFCITFNVTYLRDRTDITLIVAGPSGA 60
 DB 1 MTOAEKPFGLKRSFLMPIHTEHLKYLPMFLMFCISFNITLRDRTDITLIVAPSSGA 60
 QY 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPLRDVL 120
 DB 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPFLIFPALFTVIYPCRHIL 120
 QY 121 HPTDFADTLQAIIPSGLGLVALIRNMTFAAFYVLALMGSVMSLMFPGFANEITKISE 180
 DB 121 HPTDFADTLQAIIPSGLGLFAMLRNMTFAAFYVLSELMSGLMFMGFAEITKISE 180
 QY 181 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPMGISRLIMAMTIVSGVLMA 240
 DB 181 AKRFYALFGIGANVALISGRPAITWSKLRASLSEGVDPMGISLYFLMAMFLCSCAITAA 240
 QY 241 SYWINKNVLTDPRFYNPEEMQKKGAKPKMNKDSFLYLDSPYILLTLIVAYGIC 300
 DB 241 CYWMNRVYVLTDRFYNPAEL-KAKK-SKPKMGSEFSYLLRSPLYMLLALIVICGYC 298
 QY 301 INLEVTWKSQQLQYRNNDYSEFGNFSFTGVSVYLMFLVGNVIRKFGMLTGALV 360
 DB 299 INLEVTWKSQQLKQKQFNPNDYSAFMGNFSFTGVSVFVWMLFGVNVIRKFGMLTGALV 358
 QY 361 TPWVVLITGVFFALVIFRNQASGLVAMFGTTPMLAVVGAIONISKSTKVALFPSTK 420
 DB 359 TPWVVLITGAIIPALVIFRDHATGLVVALGTTPLMLAVVGAIONISKSTKVALFPDARK 418
 QY 421 EMAYIPLDQKQKGAIDVVAARFGSGGALIOGSLIVCGSIGAMTPYLAIVLFI 480
 DB 419 EMAYIPLDQKQKGAIDVVAARFGSGGSLIOGSLIVCGSIGAMTPYLAIVLPAII 478
 QY 481 AIMVLSATKINKLFLAQSALKEQEV-----AEDSAPAS 514
 DB 479 MWVLTSATKINKLFLAASAKKEQELAEATAAKESASPAA 518

RESULT 5

adenine nucleotide translocase, 19474-21800 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C/Accession: E96834
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hutzat, L.
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, J.; ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A/Accession: E96834
 A/Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-624 <STO>
 A:Cross-references: GB:AE005173; NID:g6751705; PIDN:AAF27687.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: PS16.5
 A/Map position: 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 50.7%; Score 1333; DB 2; Length 624;
 Best Local Similarity 51.8%; Pred. No. 2.3e-94;
 Matches 259; Conservative 89; Mismatches 140; Indels 12; Gaps 3;

QY 16 LMPIHTEHLKYLPMFLMFCITFNVTYLRDRTDITLIVAGPSGAERIPKFMVLVPCA 75
 DB 105 IFGEVATLTKIIPGLIMFPCISFNITLRDRTDITLIVAGPSAERIPKFMVLVPCA 164

Db	222	RFMGPLAIRIMISFCIFYMAELMGSVVSVLFWGFGANQITTVDEAKKFFPLFLGIGANVA	281
Qy	196	LLASGRALYMAKRLASVSEGVDPNGISLRLLMAMTIVSGVLVMASTWYIMKXNVLTDPRF	255
Db	282	LIFSRTYVYFFSNMRNLIIPGVDMGVSLKAMMSIVAGMGLAICELTWYWR-----	333
Qy	256	YNPEEMQKQKKAAPKPNMKDSFLYLDPSPIYLLTLVLVAYGICINLIEYTWKSQTLQ	315
Db	334	YVPLPLTRSKKKVYKPMQGMTEISLKLVSISPIYIDLATLVVAYGISINLVEYTWKSKLQAQ	393
Qy	316	YFNNDYSEEWGNFSEFWTGVSVLIMLFWGNAVYRKFGMLTGALVTPYVWLLTGTGVFPAI	375
Db	394	FPSPEYSASFMDDFSTCTGIAIATFWML-LSQYVFKKKGWGVAAKITPTVLLLTGVAFFSL	452
Qy	376	VIFRNQASGLVAMFGTTPMLAVVGAIONILKSTCYALFDSRTKEMAYIPLDOOKYKG	435
Db	453	ILFGEPFALVAKLGMTPLLAAYVGAIONIFSKSAKSIPLDPCKEMAYITLDEDTKYKG	512
Qy	436	KAAIDVVAARFSGSGGALLIQGGLVIGSIGAMTPYLAVILLFTIAIMLVSATKL---N	491
Db	513	KAAIDVVCNPLKSGGALLIQGFMIILFGLSLANSTPYLIGVITWMLAAKSLGQFN	572
Qy	492	KLFLLAQSALKEQF-----VAQEDSAPA	513
Db	573	TLMSSEELEREMERASSVKIPIVVSQED-APS	602

RESULT 7
 707420
 ATP/ADP-transporter, chloroplast - potato
 C:\Species\ Solanum tuberosum (potato)

R;Tjaden

A>Title: Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuberosum) tuberization

A.Reference number: Z16025

A.Accession: T07420

A>Status: preliminary; translated from GB/EMBL/DDBJ

A.Molecule type: mRNA

A.Residues: 1-631 <TJA>

A.Cross-references: EMBL:Y10821; PIDN:CAA71785.1

A.Experimental source: cv. Desiree

C.Genetics:

A.Genome: nuclear

C.Superfamily: rickettsial-type ATP/ADP translocase

C.Keywords: chloroplast; membrane protein

Query Match 49.7%; Score 1306; DB 2; Length 631;
Best Local Similarity 52.1%; Pred. No. 2,7e-92;
Matches 256; Conservative 87; Mismatches 136; Indels 12; Gaps 3;

Oy 24 LKKVLPMLFPCITFNVTVLRTKDTLLIVGAGSGEAIPFIKFWLVPCAILFMLIYA 83
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 111 LKKIIPLGMFCILENTIIRDKTVLVVTAAGSSAEIILPLKTWNLPVAIGFMILYT 170
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy 84 KLSNLISKOALFYANVGTPFLIFPALFPVTIYYPRDVLHPPREFADRLOAILPLPGLLGVAI 143
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

[illegible]

Db 403 SFGMGPSTATGATFTMML-LSQWIFDKXGMAAKITPTVLLTGVSFSLLEGAPLA 461
Qy 384 GLVAFGTTPLMLAVVGAIONILSKSTYALPDSTKEMAYIPLDQEQVYKGAIDVYA 443
Db 462 PTLAKFGMTPLAAVYVGMONIPFSKAKYSLDPCKEMAYIPLEDDTQVYKKAIDVYC 521
Qy 444 ARFGKSGGALLIOGGLLVIGSIGAMTPYLAIVLIFIAITWLASATKNTKPLAQSALKRO 503
Db 522 NPLGSGGALLIOGGLLVIGSIGAMTPYLAIVLIFIAITWLASATKNTKPLAQSALKRO 578
Qy 504 EVAQEDSAPAS 514
Db 579 EDLEKEMERAS 589

RESULT 8

S68205
ATP/ADP translocase ATP1 precursor - Arabidopsis thaliana
N/Alternate names: adenine nucleotide translocase; ATP/ADP translocator protein
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 22-Jun-1999
C/Accession: S68205
R/Kampfenkel, K.; Moehlmann, T.; Batz, O.; van Montagu, M.; Inze, D.; Neuhaus, H.E.
FEBS Lett. 374, 351-355, 1995
A/Title: Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel put
A/Reference number: S68205; MUID:96069943; PMID:7589569
A/Accession: S68205
A/Molecule type: mRNA
A/Residues: 1-589 <KAM>
A/Cross-references: EMBL:Z49227; NID:g1051108; PIDN:CA089201.1; PID:g1051109
C/Genetics:
A/Genome: nuclear
A/Suprafamily: rickettsial-type ATP/ADP translocase
C/Keywords: Chloroplast; transmembrane protein
P.1-100/Domain: translt peptide (chloroplast) #status predicted <TNP>
P.101-589/Product: ATP/ADP translocase ATP1 #status predicted <MNT>

Query Match 46.4%; Score 1220.5; DB 1; Length 589;
Best Local Similarity 51.7%; Pred. No. 8.8e-86;
Matches 252; Conservative 76; Mismatches 144; Indels 15; Gaps 4;

Qy 16 LMPITHTLKKVLPMLPFCITFNTYVLRDQTLIVAGPSSGAETPITFELVPCA 75
Db 104 IFGVAVATLKTIIPGLMEFCILFNTYVLRDQTLIVAGPSSGAETPITFELVPCA 163
Qy 76 IIFMLIVAKLSNLSKQALFYAVGTPFLIFALFPTVIYPLRDVLPHTFADRLQALIPP 135
Db 164 IGFMLIVAKLSNLSKQALFYAVGTPFLIFALFPTVIYPLRDVLPHTFADRLQALIPP 223
Qy 136 GLLGLVAILRNWTFAPFYVLAELMGSVMLSLMFWGPNETIKHAKRFPYALFGIANIS 195
Db 224 RFMGFIATIRIWSFCIFYMAELMGSVVSVLFMGPNQITVDPAKRPYALFGIANIS 283
Qy 196 LLAGSRAIVMSKLRASVSEGVDPWGISRLMAMTIYSGVLMSYTMINKVLTDRPF 255
Db 284 LIFSRATVYFSLNKKNGPVD--GSPVSHDEHCGANGTRI CUSTIGSNVYVLPFRS 341
Qy 256 YNPEEMQKKGAKPRMKNKDSFLYIDRSPIYLLTLVIAVGCINLIEVTWKSQKLQ 315
Db 342 KKKKE-----KKMGMSLKLFLVSPYIRDLATLVAVGSIINLVEVWKSQKLQ 393
Qy 316 YPNNDYSEFMGNESFMTGVSVLIMLVGNAVIRKFGMLTGALTVPVWVLLTGIVPAL 375
Db 394 FSPSMEYSAFMGAFSTCTGVATFTMML--LSQVFNKYGWGAAKITPTVLLTGVAFFSL 452
Qy 376 VIFRNQASGLVAFGTTPLMLAVVGAIONILSKSTYALPDSTKEMAYIPLDQEQVYK 435
Db 453 ILFGPFAVLAUKLMTPLAAVYVGAIONISKSAKSLPFCCKEMAYIPLEDDTQVYK 512
Qy 436 KAAIDVVAARFGKSGGALLIOGGLLVIGSIGAMTPYLAIVLIFIAITWLASATKNT 491

Db 513 KAAIDVVCNPLKSGGALLIOGGLLVIGSIGAMTPYLAIVLIFIAITWLASATKNT 572
Qy 492 KLFALQS 498
Db 573 SLRLKKS 579

RESULT 9

A97710
ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: A97710
R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: A97710
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-498 <KUR>
A/Cross-references: GB:A006914; PIDN:AAL02619.1; PID:g15619119; GSPDB:GN00173
C/Genetics:
A/Genome: tlc1

Query Match 40.9%; Score 1076; DB 2; Length 498;
Best Local Similarity 43.6%; Pred. No. 8.6e-75;
Matches 213; Conservative 103; Mismatches 153; Indels 20; Gaps 7;

Qy 11 KLRSLFPIHTHKLKVLPMFLPFCITFNTYVLRDQTLIVAGPSSGAETPITFELVPCA 70
Db 12 ELKSTVPIERENKFFLPMAFMFCILNLTSTLSTKIDGFEV--TDIGAELISFLKTYI 69
Qy 71 VPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFALFPTVIYPLRDVLPHTFADRLQ 130
Db 70 VLPASVIMNVIVYKCDILKQENVYVITSFIFGFAFVLYPPLVHPDPTISW 129
Qy 131 AILPPGLIVAILRNWTFAPFYVLAELMGSVMLSLMFWGPNETIKHAKRFPYALFGI 190
Db 130 SVAYPNVMPFRIRIVKMSFASFTYMAELMGSTMILTFWOPANQITKDEAKRFPYMGCL 189
Qy 191 GANISLASGRAIVASLRLASVSEGVDPWGISRL--LMAMTIYSGVLMSYTMINKVLTDRPF 246
Db 190 LANLALPVTYITIGCLHEKTIQI-----VAEHLKRVPLFVIMTSSFLVILTYRMKN 241
Qy 247 KNVLTDPFRFYNPEEMQKKGAKPRMKNKDSFLYIDRSPIYLLTLVIAVGCINLIEV 306
Db 242 KNVLTDPRLYDPAVYK--EKAKAKMSLIDSPKMLFTSKYGYIALLLIAYGVSVNLVWG 299
Qy 307 TWKSQKLQYPNNDYSEFMGNESFMTGVSVLIMLVGNAVIRKFGMLTGALTVPVWV 366
Db 300 VKMSKYKELVPTKEAYTYTMKGFQYQGVAVIAEVL--IGSNILRRVSWLTAMITPLMML 358
Qy 367 LTVGFELVIFRNQ--ASGLVAMGTPPLMLAVVGAIONILSKSTYALPDSTKEMAYI 425
Db 359 ITGAAPFAFIFPDSYIAHLTGILASGPLALAVMIGMIONVLSKGVKSLDPATONMYI 418
Qy 426 PLDQEQVYKGAIDVVAARFGKSGGALLIOGGLLVIGSIGAM--TPYLAIVLIFIAITW 483
Db 419 PLDQEQVYKGAIDVVAARFGKSGGALLIOGGLLVIGSIGAM--TPYLAIVLIFIAITW 478
Qy 484 LVSATKLNK 492
Db 479 IYAVGGLNK 487

RESULT 10

U00026
ATP/ADP translocase tlc1 - Rickettsia prowazekii
N/Alternate names: ADP/ATP carrier protein tlc1; RP053
C/Species: Rickettsia prowazekii
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 03-Nov-2000

C:Accession: J00026; E71713
R:Williamson, L.R.; Plano, G.V.; Winkler, H.H.; Krause, D.C.; Wood, D.O.
Gene 80, 269-278, 1989
A:Title: Nucleotide sequence of the *Rickettsia prowazekii* ATP/ADP translocase-encoding
A:Reference number: J00026; MUID:90060776; PMID:2555259
A:Accession: J00026
A:Molecule type: DNA
A:Residues: 1-498 <NT>
A:Cross-references: GB:M2816; NID:9152469; PIDN:AAA6382.1; PID:9152470
R:Anderson, S.G.E.; Zomrodipou, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: E71713
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-498 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA14524.1; PID:9386062
A:Experimental source: strain Madrid E
A:Comment: This ATP/ADP translocase shares no detectable amino acid sequence homologies
C:Genetics:
A:Gene: tlc1; RP053
C:Function:
A:Description: exchanges ADP in the rickettsial cell with ATP in the host cell
C:Superfamily: rickettsial-type ATP/ADP translocase
C:Keywords: transmembrane protein
F:28-45/Domain: transmembrane #status predicted <TR01>
F:68-84/Domain: transmembrane #status predicted <TR02>
F:93-115/Domain: transmembrane #status predicted <TR03>
F:183-206/Domain: transmembrane #status predicted <TR04>
F:219-237/Domain: transmembrane #status predicted <TR05>
F:279-297/Domain: transmembrane #status predicted <TR06>
F:346-370/Domain: transmembrane #status predicted <TR07>
F:380-399/Domain: transmembrane #status predicted <TR08>
F:445-461/Domain: transmembrane #status predicted <TR09>
F:466-482/Domain: transmembrane #status predicted <TR10>

Query Match 40.2%; Score 1056; DB 1; Length 498;
Best Local Similarity 41.6%; Pred. No. 2.9e-73;
Matches 207; Conservative 106; Mismatches 165; Indels 20; Gaps 7;

QY 2 TTKTEKPEKGRKSPFMPPIHTHEKLVLPWLFPCITFNTYVLRDTKTLIVAGAGSGAE 61
DB 3 TSKSENIVSELRKIIWPIEQYENKKFLPLAFMPCILNYSLTRSIKSGFV--IDIGTE 60

QY 62 AIPPIKWLVPQCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPALFPTVYIPLRDVLH 121
DB 61 SISFLKTYIVLPSAVIAMIIYVKLCIDIKQENVFYVITSFFLGYPALFAFVLYPYPDVLH 120

QY 122 PTEFADRLQALIPGILGIVAILRNWTFPAFYVLAELMGSVMLSLMFGFANEIKIHEA 181
DB 121 PDKHTISLAVYPMFKFKIKIVGRKSPFYTIALMGTWMLSLFVQFAQITKIAEA 180

QY 182 KRFVALFGIGANISILASGRAIWMASKLRASVSEGVDPWGISLR-----LMAMTIVSGIV 237
DB 181 KRFYMPGILNIALPVSIVTGYFLHETQI-----VAHLKFPLEFVIMTSSFL 232

QY 238 LMASTWINKNVLTDPFRFYNPEEMQKKGAKPRKMMKDSFLYLDSPYIILLTLITAY 297
DB 233 ILITRYMNNKIVLTDPRLVDPALVK--EKTKAKLSFISLKMIFTSKVVGIALLITAY 290

QY 298 GICINLIEVTWSQKLQYPMNDYSEPMGNFSFMTGVSVLIMFVGGNVIRKSGWLTG 357
DB 291 GVSVALVGVGWSKVKELPTREAYTIVGQFQGVWALAIFM--IGSNILRKVSWLTA 349

QY 358 ALVTPVMTLTGIVFPFALVIFRNO--ASGLVAMFGTPTLMAVVGALQNIILSKSTRYALF 416
DB 350 AMITPLMFPITGAAPFSFIFPDSVIAMLTGLASSPLTILAMIGMIVLSKGYKISLF 409

QY 417 DSTKEMAYIPLDQOKVKGKAAIDVVAARFGKSGALIQOGLLVI CGSIGAM--TPYLAIV 474
DB 410 DATKMAVYIPLDKDRLVRGQAIVAVIYIGRLGKSGALIQSTFIFLPVFGFIEATPYPAS 469

QY 475 ILFFIATMVSATKANK 492
DB 470 IFFIIVIMIRAVKANK 487

RESULT 11
ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: C86567
R:Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: C86567
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-540 <STO>
A:Cross-references: GB:BA000008; NID:98978986; PIDN:BA98821.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: act 2
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 39.0%; Score 1025.5; DB 2; Length 540;
Best Local Similarity 39.4%; Pred. No. 6.9e-71;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

QY 1 MTKTEKPEKGRKSPFMPPIHTHEKLVLPWLFPCITFNTYVLRDTKTLIVAGAGSGA 60
DB 1 MOSSVVRKFSRLRAVLCIYKSEFSKFPFLAFVGVNCLKNMMDLIVIGSDAGA 60

QY 61 EAIPIKWLVPQCAIIFMLIYAKLSNLSKQALFYAVGTPLIFPALFPTVYIPLRDVL 120
DB 61 EVIPIKWKGVPGAVIYVTVYGMVLSRYPDITFYCFMAAFLGFFFAVLIYVPGDSL 120

QY 121 HPTFADRLQALIPGILGIVAILRNWTFPAFYVLAELMGSVMLSLMFGFANEIKIHE 180
DB 121 HLNSLADLQQLLPQGLRGFIWMRYMSYSIYVMSLSVSLMPLGMLAQITITTE 180

QY 181 AKRFVALFGIGANISILASGRAIWMASKLR--ASVSEGVDPW--GISRLMLAMTIVSGIVL 238
DB 181 AGRFVALINTGLNLSICAGISYWMKQTFVAASPADSHSVNLNLTMLIT--CSGLIM 239

QY 239 MASTWINKNV--LTPFRFYNPEEMQKKGA-----KPKMMKDSFLYLDSPYI 287
DB 240 I-----MLYRRIHMLTIDTISIPSRVLAEEGAATANKKKKPKAKANLFLHLIQRYL 295

QY 288 LLTLTVAAGICINLIEVTWSQKLQYPMNDYSEPMGNFSFMTGVSVLIMFVGGN 347
DB 286 LGLATIVSNLVHLEFVWKQDQSQIYSSHVEENGMSRITLIGVSVLAVALLTGQ 355

QY 348 VIRKGMVLGALVTPVMTLTGIVFPFALVI--FRNOASGLVAMFGTPTLMAVVGALQNI 406
DB 356 CIRKGMVVGALVTPVMTLVSGLLFFGTIFPAKQDISIFGVLTMTPALAAMGGMQNV 415

QY 407 LSKSTKALFSTYEMAYIPLDQOKVKGKAAIDVVAARFGKSGALIQOGLLVI CGSIG 466
DB 416 LSRGKFFPQOTKEMAFIPLSPEDKNHGKAAIDGVVSRICKSGSLIQOGLVIFSSVA 475

QY 467 AMTYLVAILLFIATMVSATKANKLFLAQA-----LK-----BOEVQOE 508
DB 476 ASLVIALVLLIIMVMAIVAVIYIGKEYSRADAVALTKPKPSSSVIREADESVQOE 535

QY 509 DSA 511
DB 536 EMA 538

RESULT 12
B72056
ADP, ATP carrier protein CP0133 [imported] - Chlamydia pneumoniae (strains CWL029 an

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: B72056; G81610
 R:Katman, S.; Mitchell, W.; Marache, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: B72056
 A:Molecule type: DNA
 A:Residues: 1-540 <ARN>
 A:Cross-references: GB:AE001646; GB:AE001363; NID:g4376910; PIDN:AD18753.1; PID:g437692
 A:Experimental source: strain CML029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: G81610
 A:Molecule type: DNA
 A:Residues: 1-540 <REA>
 A:Cross-references: GB:AE002174; GB:AE002161; NID:g7189059; PIDN:AA838016.1; PID:g718906
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: act 2; CP0133
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 39.0%; Score 1025.5; DB 2; Length 540;
 Best Local Similarity 39.4%; Pred. No. 6.9e-71;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

QY 1 MTKEERKFGKLRSLPFIHTEHLKVLPMFLMFCITFENYTVLRDTKDTLIVAGPSGA 60
 1 MOSSEVKPFSRLKALCYKSEPSKPYPLFLAFVGFNYCLKNKMDLIVIGSDAGA 60
 DB 61 EALPFIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPYIYPLRDL 120
 61 EVIPFLKVGIVGAVIYVWYGMVGRYPRDYFYCFMAFLGPFPLFAVIITIPVGSGL 120
 QY 121 HPTFADRLQALIPGLIGLVAIRNTPFAAFYLAELMGVMSLMFWGPAEITKHE 180
 121 HNSLADQLQELPGLQGLFIVWVWYSIYVWSELMSVLSMLFWGLANOITITTE 180
 DB 121 HNSLADQLQELPGLQGLFIVWVWYSIYVWSELMSVLSMLFWGLANOITITTE 180
 QY 181 AKRFYALFGIGANISLASGRAIYVASKLR-ASVSEGVDPW-GISRLMTMTVSGVL 238
 181 AGRFYALINLNLMSISICAGEISTYMGKQTVAAISFACDSWHSVMLNLTMLIT-CSGLIM 239
 DB 181 AGRFYALINLNLMSISICAGEISTYMGKQTVAAISFACDSWHSVMLNLTMLIT-CSGLIM 239
 QY 239 MASYWINKNV-LTDPFYNPBEKQKGA-----KPKNMKDSFLYIDRSPYI 287
 240 I-----MIVRRIMHLLTIDTSPSRVLAEEGAATANLKEKKKPKAKANLFLHLIQSRYL 295
 DB 240 I-----MIVRRIMHLLTIDTSPSRVLAEEGAATANLKEKKKPKAKANLFLHLIQSRYL 295
 QY 288 LLLTLVLVAYGICINLIVTWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGN 347
 288 LLLTLVLVAYGICINLIVTWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGN 347
 DB 288 LLLTLVLVAYGICINLIVTWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGN 347
 QY 296 LGLAILIVLSYMLVHLFEVWKDQVSQIYSHVENGYMSRITTLIGVSVLAVALTLTGQ 355
 296 LGLAILIVLSYMLVHLFEVWKDQVSQIYSHVENGYMSRITTLIGVSVLAVALTLTGQ 355
 DB 296 LGLAILIVLSYMLVHLFEVWKDQVSQIYSHVENGYMSRITTLIGVSVLAVALTLTGQ 355
 QY 348 VIRKFGMTGALVTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONI 406
 348 VIRKFGMTGALVTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONI 406
 DB 348 VIRKFGMTGALVTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONI 406
 QY 356 CIRKMGMTVGLVLPVWLVSGLLFGTTPAKKADISLFGGVLLPALTALAMTGQMN 415
 356 CIRKMGMTVGLVLPVWLVSGLLFGTTPAKKADISLFGGVLLPALTALAMTGQMN 415
 DB 356 CIRKMGMTVGLVLPVWLVSGLLFGTTPAKKADISLFGGVLLPALTALAMTGQMN 415
 QY 407 LSKSTKVALPDSTKEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIVICSGIG 466
 407 LSKSTKVALPDSTKEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIVICSGIG 466
 DB 407 LSKSTKVALPDSTKEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIVICSGIG 466
 QY 416 LSRGTGKTFPQOTKEMAFIPLSPEDKXNGKALIGVSRICKSGSGLIYQGLVIFSSVA 475
 416 LSRGTGKTFPQOTKEMAFIPLSPEDKXNGKALIGVSRICKSGSGLIYQGLVIFSSVA 475
 DB 416 LSRGTGKTFPQOTKEMAFIPLSPEDKXNGKALIGVSRICKSGSGLIYQGLVIFSSVA 475
 QY 467 AMPTYVALVILFLTAIWLVSATKLNKFLAOSA-----LK-----EDEVAOE 508
 467 AMPTYVALVILFLTAIWLVSATKLNKFLAOSA-----LK-----EDEVAOE 508
 DB 467 AMPTYVALVILFLTAIWLVSATKLNKFLAOSA-----LK-----EDEVAOE 508
 QY 509 DSA 511
 509 DSA 511
 DB 509 DSA 511
 QY 536 EMA 538
 536 EMA 538
 DB 536 EMA 538

ADP, ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C:Accession: B97790
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: B97790
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-501 <KRR>
 A:Cross-references: GB:AE006914; PIDN:ALU03260.1; PID:g15619815; GSPDB:GN00173
 C:Genetics:
 A:Gene: rlc3
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 37.1%; Score 976.5; DB 2; Length 501;
 Best Local Similarity 40.0%; Pred. No. 3.6e-67;
 Matches 199; Conservative 106; Mismatches 174; Indels 19; Gaps 8;

QY 9 FGKRSFLMPFIHTEHLKVLPMFLMFCITFENYTVLRDTKDTLIVAGPSGAELPIKE 68
 8 FEKVKELMPEIRKELKLFIPMALMCLIFNFGALRSIDSLV--PSMGAEIISFLK 65
 DB 69 MIVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPYIYPLRDLRPT-EPAD 127
 69 MIVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPYIYPLRDLRPT-EPAD 127
 QY 69 MIVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPYIYPLRDLRPT-EPAD 127
 DB 69 MIVVPCALIFMLIYAKLSNLSKQALFYAVGTPELLIFALFPYIYPLRDLRPT-EPAD 127
 QY 128 RLQALIPGLIGLVAIRNTPFAAFYLAELMGVMSLMFWGPAEITKHEARFYAL 187
 128 RLQALIPGLIGLVAIRNTPFAAFYLAELMGVMSLMFWGPAEITKHEARFYAL 187
 DB 128 RLQALIPGLIGLVAIRNTPFAAFYLAELMGVMSLMFWGPAEITKHEARFYAL 187
 QY 126 KLIASV-PNKRWFIKISQMSYALMTYFALMSAVVNTLMFQPAHNIPTSKAKRPYV 184
 126 KLIASV-PNKRWFIKISQMSYALMTYFALMSAVVNTLMFQPAHNIPTSKAKRPYV 184
 DB 126 KLIASV-PNKRWFIKISQMSYALMTYFALMSAVVNTLMFQPAHNIPTSKAKRPYV 184
 QY 188 FGIGANISLASGRAIYVASKLRASVSEGVDP-----WGISRLMTMTVSGVL 239
 188 FGIGANISLASGRAIYVASKLRASVSEGVDP-----WGISRLMTMTVSGVL 239
 DB 188 FGIGANISLASGRAIYVASKLRASVSEGVDP-----WGISRLMTMTVSGVL 239
 QY 185 LGWGNIGLIIAGSVLVFPSSGQVDISELLPDSFNSAGAHMLQPMISIIYVAGIIM 244
 185 LGWGNIGLIIAGSVLVFPSSGQVDISELLPDSFNSAGAHMLQPMISIIYVAGIIM 244
 DB 185 LGWGNIGLIIAGSVLVFPSSGQVDISELLPDSFNSAGAHMLQPMISIIYVAGIIM 244
 QY 240 ASYWINKNVLTDPFRFNPBEKQKGAAPKAMKDSFLYIDRSPYIITLLIVAYGI 299
 240 ASYWINKNVLTDPFRFNPBEKQKGAAPKAMKDSFLYIDRSPYIITLLIVAYGI 299
 DB 240 ASYWINKNVLTDPFRFNPBEKQKGAAPKAMKDSFLYIDRSPYIITLLIVAYGI 299
 QY 245 LFPFIIRFLITDS--INVLDAKKVTAQMTKLSVIESIKLVHSHKYGRIALLIITYGL 302
 245 LFPFIIRFLITDS--INVLDAKKVTAQMTKLSVIESIKLVHSHKYGRIALLIITYGL 302
 DB 245 LFPFIIRFLITDS--INVLDAKKVTAQMTKLSVIESIKLVHSHKYGRIALLIITYGL 302
 QY 300 CINLIEVWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGNVIRKFGMTGAL 359
 300 CINLIEVWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGNVIRKFGMTGAL 359
 DB 300 CINLIEVWKSQKLOYPNMNDYSEPMGNFSFWTGVSVLIMLVFVGNVIRKFGMTGAL 359
 QY 303 LNIIVEGPMKAKIKELHPNITIDYVNGRINMGISCVPFI-IGSNIIRLGLMSAL 361
 303 LNIIVEGPMKAKIKELHPNITIDYVNGRINMGISCVPFI-IGSNIIRLGLMSAL 361
 DB 303 LNIIVEGPMKAKIKELHPNITIDYVNGRINMGISCVPFI-IGSNIIRLGLMSAL 361
 QY 360 VTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONIILSKSTKYPALPDST 419
 360 VTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONIILSKSTKYPALPDST 419
 DB 360 VTPVWVLTLGIVFALVI-FRNOAGSLVMPFGTTPMLLVVGAIONIILSKSTKYPALPDST 419
 QY 362 LTPIMLSITGLMFPFIIFIEICEGCGDFNL-LVAAILVGAIONILSKSKSLPDS 419
 362 LTPIMLSITGLMFPFIIFIEICEGCGDFNL-LVAAILVGAIONILSKSKSLPDS 419
 DB 362 LTPIMLSITGLMFPFIIFIEICEGCGDFNL-LVAAILVGAIONILSKSKSLPDS 419
 QY 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 DB 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 QY 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 DB 420 KEMAYIPLDOEKVKGKAIIDVVAARFGSGGALLIOGGLIV--CGSIGMTPLAVIIL 477
 QY 478 FIATIMLVASATKLNKFL 495
 478 FIATIMLVASATKLNKFL 495
 DB 478 FIATIMLVASATKLNKFL 495
 QY 480 VMMSLMIMVYIKANKERV 497
 480 VMMSLMIMVYIKANKERV 497
 DB 480 VMMSLMIMVYIKANKERV 497

RESULT 13 B97790

A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: tlc3; RP477
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.9%; Score 969.5; DB 2; Length 501;
Best Local Similarity 39.6%; Pred. No. 1.2e-66;
Matches 197; Conservative 107; Mismatches 176; Indels 17; Gaps 7;

Oy 9 FGKLRSPMPHITHELKVLPMFMFCITFNTYVLRDTKTLIVGAGSGAEALPIK 68
Db 8 FEKVEIIMPIERKELKFIIPALMMTCILFNFGALRSIKDSLVV--PSMGAEIISFLKL 65
Oy 69 WLWVCAIIFMLIVAKLSNLSKQALFYAVGTFPIIFALPFTVYPLRDVLPHTPADR 128
Db 66 WLWVSCVITFIIYKLSNKLNFETIFYSIVGTFILFELPFIYIIPMODIHPDAMIN 125
Oy 129 LQALIPGLGLVALIRNMTFAAFVYLAELMGVSLMFMGFANEITKIHAKRFVALF 188
Db 126 NLIASYPMUKMFIRKIGSKSYALMTIFSELMSAVINIMFQFANHIFDTAKAKRFYVL 185
Oy 189 GIGANISLASGRAIVMSKLRASVSEGV--DPM-----GISRLMAAMTIVSGLVMA 240
Db 186 GMGVNGIIGIINGSVLFFSSGGYIIDSBLTDSYSSNNNSIMLOPIIISITVTAIGIMF 245
Oy 241 SYWINKVAVLDPFRYNEEMOKGKAKPKMMKDSFLYDRSPYIILLTLIVAGIGIC 300
Db 246 LFRINKITLINS--INVLDYKVAAKTKTKALIESIKLIHSKTYIGRIALLIICYLL 303
Oy 301 INLIEVTWKSQKLOYPNNNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKFGMLTGLV 360
Db 304 INIVGPKAKIKELHPPTVDYVFMGMFNIMMGISCTPMI--IGSNITRLRLGMLISALL 362
Oy 361 TPVAVLTLGIYFPALVIRNOASGLVAMFGTTPMLAVVGAIONILSKSKTKYALFSDTK 420
Db 363 TPIMLSITGFMPFIIFIEIEIGTCFGDENV--LVVAIIIVGAIQNIILSKSKYSFLFSDTK 420
Oy 421 EMAYIPLDQEQKKAIDVVAARFGSGGALIQGGLV--CGSIGAMPYLAIVLILF 478
Db 421 EMAYIPLELRTKKAIVEYIGTKFGKLGAFIOSLFIITIPATPDSIITLIVLIV 480
Oy 479 IIAIMLVASATLKNLFL 495
Db 481 MAMLMIMWIIITLKNKEYI 497

RESULT 15

probable adp/atp translocase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71503
R:Stephens, R.S.; Kalman, S.; Jammal, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: E71503
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-540 <ARN>
A:Cross-references: GB:AE001323; GB:AE002173; NID:g7328931; PIDN:AA68096.1; PID:g332893
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT495
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.4%; Score 956.5; DB 2; Length 540;
Best Local Similarity 37.2%; Pred. No. 1.3e-65;
Matches 197; Conservative 116; Mismatches 193; Indels 23; Gaps 7;

Oy 4 TEKRPFGKLSFLMPHITHELKVLPMFMFCITFNTYVLRDTKTLIVGAPSGAEAI 63
Db 3 SEVKSFSKRGYFPPIYRSFSPKFIPLFLAFVGVNVALKTKTDSLIVLVSRAAEVI 62

Oy 64 PFIKFWLVPICAIIFMLIVAKLSNLSKQALFYAVGTFPIIFALPFTVYPLRDVLPHT 123
Db 63 PFLKVGIVGAVAVIMTYGMSRRYSRGVIFSLVGGFLGFPALFPIYIIPIDALHLN 122
Oy 124 EPADRLOALIPGLGLVALIRNMTFAAFVYLAELMGVSLMFMGFANEITKIHAKR 183
Db 123 KLAAGKQIILPPGGRGVVWQVWSYSLYVMSLMSIVSLTFWGVANHITSVREAGR 182
Oy 184 FYALFGIGANISLASGRAIVMSKLRASVSEGVDM--GISRLMAAMTIVSGLVMA 241
Db 183 FYALINIGLMSVFAEBSVLIGRNPVIAFPVADWHEVLNITLIIYLAGGVIL--- 239
Oy 242 YWINKVAVLDPFRYNE-----EMOKGKAKPKMMKDSFLYDRSPYIILLTL 292
Db 240 YLVKDLRLDETSMLKEGLAEBSVQKKEKRSQAK--AKSLPALLRSRLIGAV 297
Oy 293 LVYAGICINLIEVTWKSQKLOYPNNNDYSEFMGNFSFWTGVSVLIMLVGNAVIRKF 352
Db 298 VLISYNLVIHLFEVVMKDQCRVYASRVERFSVMSRITTLTGIVSALTGFPAAGOTIRW 357
Oy 353 GMLTGAIVTPVMTLTIIVFF--ALVIRNOASGLVAMFGTTPMLAVVGAIONILSKST 411
Db 358 GMTIGALVPLTMLITGALFFGAIYAVGDMITFGILGISPLVTAMLGQVQVFSRAI 417
Oy 412 KYALFSDTKEMAYIPLDQEQKKAIDVVAARFGSGGALIQGGLVIGSIGATPY 471
Db 418 KFTTFDQTKEMAFPLEDEDEKNGKKAIDGVISRVGSGSLVYQGLIIFSSVAASLNA 477
Oy 472 LAVILFIIAIVASATLKNLFLAOGA-----LKEQEAQOESDAPAS 514
Db 478 ITVILALISGWFVIAIMLGREYAKTEALFRVNVSEDLQEBREASS 526

RESULT 16

ADP, ATP carrier protein TC0782 [imported] - Chlamydia muridarum (strain N19S)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81665
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: F81665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-543 <RET>
A:Cross-references: GB:AE002346; GB:AE002160; NID:g7190805; PIDN:AAF39585.1; PID:g719080
A:Experimental source: strain N19S (MoPn)
C:Genetics:
A:Gene: TC0782
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 36.2%; Score 953; DB 2; Length 543;
Best Local Similarity 37.5%; Pred. No. 2.5e-65;
Matches 199; Conservative 111; Mismatches 197; Indels 24; Gaps 7;

Oy 4 TEKRPFGKLSFLMPHITHELKVLPMFMFCITFNTYVLRDTKTLIVGAPSGAEAI 63
Db 3 SEVTFKFRFRYPPIHKSEFPKFIPLLLAFVGFVNSLTKTKDSLIVLAGSRAAEVI 62
Oy 64 PFIKFWLVPICAIIFMLIVAKLSNLSKQALFYAVGTFPIIFALPFTVYPLRDVLPHT 123
Db 63 PFLKVGIVGAVIITMITYGMSCRYSRGVFCALVGGFSFPAALFACVITYPMGDALHLN 122
Oy 124 EPADRLOALIPGLGLVALIRNMTFAAFVYLAELMGVSLMFMGFANEITKIHAKR 183
Db 123 GLAAKQIILPPGGRGVVWQVWSYSLYVMSLMSIVSLTFWGVANHITSVREAGR 182
Oy 184 FYALFGIGANISLASGRAIVMSK--LRASVSEGVDM--GISRLMAAMTIVSGLVMA 241
Db 183 FYALINIGLMSVFAEBSVLIGRNPVIAFPVADWHEVLNITLIIYLAGGVIL--- 239

Db 183 FYALINVLGIVNAGSIVAGEISLWLGKHTLIPSSMAVDAMHGVLLNTLLIIVAGGLL- 239
Qy 242 YMMINKNTLTD- - - - - RFYNEEMQKGGKAKKPKMMKDSFLYIDRPPYLLTL 292
Db 240 YYYRKLIDHTEBAPVLDGLVSEMSVAQKQK- - RPKAKAKSLSLVYFRSRYLMLGAV 297
Qy 233 IIVAYGICINLIEVTWKSQTLQYPMNDYSEPMGNFSFMTGVSVLYLMPVGNVIRKF 352
Db 238 VVLAINVLAHLEAVWKEVQCIYSRVEFNSYMSRITFTGIVSAVLGAVFAAGSIRRW 357
Qy 353 GMLTALVTPVAVLLTGIVF-ALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411
Db 358 GWTVALITPLTMLITGLFPGAIYAVKGDAMILGGLFSPFLVTLAMLGIVQVNFSPAI 417
Qy 412 KYALDSTKEMAYIPLDQOKYKGRALIVVAPRKSGLALIOGLIVICSGISGAMPY 471
Db 418 KTYPDQTEKAFIPLEDEKQYKRAIDGIVSRKSGSGLVYALLLIFSSVADCMNA 477
Qy 472 LAVILFIIAYMLVSAFKLKLFLAQS- - - - - ALKEQEVNAQEDSAPASS 515
Db 478 ITIVILLALGWIWVAVMLGKYSVTRTALGKRAAEPSLDEDESRVSS 528

RESULT 17

B97842
ADP, ATP carrier protein (imported) - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: B97842
R:Ogata, H.; Audic, S.; Renezo-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: B97842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-499 <KUR>
A:Cross-references: GB:AB006914; PIDN:AAL03676.1; PID:g15620264; GSPDB:GN00173
A:Gene: C1G5
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 34.5%; Score 907.5; DB 2; Length 499;
Best local similarity 37.4%; Pred. No. 6,9e-62;
Matches 190; Conservative 107; Mismatches 176; Indels 35; Gaps 13;
Qy 1 MTKTEKDF-GKLRSFLPIHTEHLKVLPMFLMFCITFNTYVLRDTKDTLIVAGPSG 59
Db 1 MLSTSRSPKONKFRAPFVNVHVELGKFIPISTLMFCILFNONVLRILKDSILISE--IS 58
Qy 60 AEAIFPIKFWLVPCALIFMLIYAKLSNLSKQALFYAVGTEPFLIPALFPYIYPLADV 119
Db 59 AEAIAFAKYCVTPAALFVLIYAKMINVLTPEKIFYVLSAFISFVLFTEVIYPINIH 118
Qy 120 --LHPTFADRLQALIPGLGLVALIRNWTFAFYVLAELMGSVMLSMFGFANEITK 177
Db 119 FTHVHNMLADWERY--PHEKMYISLVGNGWGIYVLSAELMPLNFYVLLFQFANEILT 176
Qy 178 IHEAKRFYALFGIGANISLASGRAIWNAS- - - - - KLRASVSEG-VDPMWISLRLMAM 230
Db 177 TBEAKRFYTLFSLFQNSSLILVGFPMNLSSEDTIHKFMSISDSKITLVQSTTIVAI 236
Qy 231 TIVSGVLMASTWYNKQVLDPRRYNPEMOKGKG- - AKPKNMKDSFLYLDSPYIL 288
Db 237 AIIICLLVR- - - - - FISKVFTNPLFY- - - - - AKASGRSTSERMGLIKSFYIAKSKYLW 287
Qy 289 LITLIVAYGICINLIEVTWKSQTLQYPMNDYSEPMGNFSFMTGVSVLYLMPVGNV 348
Db 288 LILISGAIFGPAINLVEAVWKAKIKELIPTVNTVIEFNSLYILMG-VAIHWMTIIGNNT 346
Qy 349 IIRFGMLGALVTPYVAVLLTGIVFALVIFRQA- - - - - SGLVAMFGTTPMLAVVGAIO 404
Db 347 MRMHMFVAAVISPIYIMVTGILFVLLVIFDQIISLFDGAILM--SPLALAVSIGIO 403

Qy 405 NILSKTYALFDSKEMAYIPLDQOKYKRAIDVVAARFGSGALIOGL--LYIC 462
Db 404 NILAKGYSTWDSREMLYPLDEBLTKGAADVIAKVGSSSGVOSITFTLVPT 463
Qy 463 GSIGAMTPYLAIVLFIAYMLVSAFKL 490
Db 464 ATFTLISPLWVFFVFLAMIVAVRKI 491

RESULT 18

B97783
ADP, ATP carrier protein (imported) - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: B97783
R:Ogata, H.; Audic, S.; Renezo-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: B97783
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <KUR>
A:Cross-references: GB:AB006914; PIDN:AAL03204.1; PID:g15619755; GSPDB:GN00173
A:Gene: C1G4
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 34.5%; Score 907.5; DB 2; Length 511;
Best local similarity 37.9%; Pred. No. 7.1e-62;
Matches 187; Conservative 111; Mismatches 175; Indels 21; Gaps 9;
Qy 9 FGKRSFLPIHTEHLKVLPMFLMFCITFNTYVLRDTKDTLIVAGPSGALPIKIF 68
Db 19 FSKLIDYMPIKRHSKFLFTLIMFCILFIQNLIRALKOSIVTM--IGAEIISFLKF 76
Qy 69 WLVPICALIFMLIYAKLSNLSKQALFYAVGTEPFLIPALFPYIYPLADVH--PTEFA 126
Db 77 WGVPSAFMLTAIVKLVNKKAKENIFLIISIFLTFPALFAVYIFPHMELHSPVTVQ 136
Qy 127 DRLQALIPGLGLVALIRNWTFAFYVLAELMGSVMLSMFGFANEITKHEAKRFA 186
Db 137 NLMSL--PNLKWPIWLSKSPSFLYIABLMPVAVALLFMQFVNNTTVEBSKREYP 194
Qy 187 LFGIGANISLASGRAIWNASKLRASVSEGVDPWG- - - - - ISLRIMAMTIVSGVLMA 242
Db 195 LFGILSQGIVLAGQFLENLSNINDYVTKRPAQSSFTLSIQILITVLIGIAIKTF 254
Qy 243 WINKNVLTDRFPYNPEMOKG-KGAKPKNMKDSFLYLDSPYIILLTLIVAYGIC 301
Db 245 WLNHKKVL- - - - - DKEMALLRFRKAKKSKWTIASEFMLLSRIIRIATLLCYGLAI 308
Qy 302 NLEIETWKSQTLQYPMNDYSEPMGNFSFMTGVSVLYLMPVGNVIRKFGMLTALVT 361
Db 309 NLVSGPMFAATKIKYKPTFPAALFISLSTYGFTTILFVV-LGSNIYRKLMPFAVIT 367
Qy 362 PVMVLTGIVFALVIFRQASGLVAMF-GTTPMLAVVGAIONILSKSTKTYALFDSK 420
Db 368 PLVITGILFFAVNNFRPAGLIIANLIDPALIATIGALQNVLSKSKYTLFDSK 427
Qy 421 EMAYIPLDQOKYKGAIDVVAARFGSGALIOGLIVICSGISGAMTPYLAIVLFI 480
Db 428 EMAYVPLDPEIKIKGKAADVIGTKGSGSGLFQSLVFIILPSASYOSTICIMIFI 487
Qy 481 A--IWLVSATKLAK 492
Db 488 TGLTLMATKALNK 501

RESULT 19

F71563
ADP, ATP carrier protein (C1G4) RP500 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: F71633
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:982893
A:Accession: F71653
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA14952.1; PID:g386105
A:Gene: tlc4; RP500
C:Superfamily: rickettsial-type ATP/ADP translocase .

Query Match 33.5%; Score 880.5; DB 2; Length 512;
Best Local Similarity 36.9%; Pred. No. 8.3e-60;
Matches 184; Conservative 108; Mismatches 177; Indels 29; Gaps 10;

QY 9 FGKRSFLMPPIHTEHKLVKLPMLFPCITPNYTVLRDXTDLIVGAGSGAARPIK 68
DB 19 FSKLTDYIWPICKRHISKFLFTLLMFCILQNLRLKOSIVTWM--IGAEITSLKF 76
QY 69 WLVPFCALIFMLIYAKSNILSKQALFPAVGTPELIFALFPYIYPLRDYLH--PLEFA 126
DB 77 WGVNPSAFLIVIVYKLVNRMKAENIFYLIIISLTFEPALFAVYIFPHEMLHLPVT-V 135
QY 127 DRLOALDPGLGLGVALIARNTFPAFYVLAEIAGSVMLSLMFQFANEITKHEAKDEYA 186
DB 136 HNLTRSL-PNLKWFLLLSKMSFSLFTIILAEMLPVVAFALFPQVANNITVESKRFYP 194
QY 187 LFGIGANISLISAGRAIYVASKLRASVEGVDPMG---ISLRLLMNTIVSGVLNMSY 242
DB 195 LFGILSQGIYLAHGFLENLSINIVYTNKRALQSPHTTISIQIILTLIVLGIVSKTF 254
QY 243 WINKNVLTDP----RTPNPEBNQKKGKAKPKMNKDSGLTYDRSPYIILLTLVIAY 297
DB 255 WLNHRKVLDDKGMALLRP-----KTRNKSITIAKSPQMLSSRHRLTLTLICY 304
QY 298 GICINLIEVTWKSQOLKQYPMNNDYSEPMGNFSFMTGVSVLIMLPVGNVIRKFGMLTG 357
DB 305 GIALLNVEGPKKAATKTKYKTPTEYAPLTSYLSYIGVFITLFFVL-LGSNIVRMGFST 363
QY 358 ALVTPVWMLLTGIVFALVIFRNQASGLVAAF-GTTPPLMLAVVGAIONILSKSTKALF 416
DB 364 AVITPSIVITIGILFPAANNFEGFAGLIIFILDPALVALITIGALQNTLSKSKXTLFF 423
QY 417 DSTKEMAVIPLDQOKVKGKAAIDVVAARFGSGGALIQGLLVICSGISGAMTPYLAAIL 476
DB 424 DSTKEMAVVPLEPEIKTISGKAADVTGFKLGSAGSAPFQSLHFIILPSASYSISICMI 483
QY 477 LFIIL-ALVWLSATKLNK 492
DB 484 IFILTCVTWATKELNK 501

RESULT 20
G71633
ADP/ATP carrier protein (tlc5) RP739 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: G71633
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:982893
A:Accession: G71633
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-500 <AND>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15167.1; PID:g386124

A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: tlc5; RP739
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 32.9%; Score 965; DB 2; Length 500;
Best Local Similarity 35.7%; Pred. No. 1,2e-58;
Matches 176; Conservative 109; Mismatches 182; Indels 26; Gaps 10;

QY 11 KLSFLPMTHTHEIKVLPMFLMFCITENVYTLVDRDTKTLIVAGPQSGAEALPIFKWL 70
DB 13 KFRFAFPVHNVELGKFIPIISALMFCITLPMQNLIRLIKOSILISE--ISABIGFAKVCY 70
QY 71 VVPCALIFMLIYAALNSILSKQALEVAGNPLIFLPAFTVLYPLRDV--LHPTPADR 128
DB 71 VTPAALFVLIYAAMINHLTFEKEI FYLSAFESCFLFPAFYVYPNHIHVPDPLSDW 130
QY 129 LQALIPGLGLVALINMTFPAAPVYLAELMGSVMTSLMFWGFANBITIKHEAKRYALP 168
DB 131 MNKY--PHEFKYISLVGNMGYIYYISLAELMPNIFVYLFMQTNELTTEAKRFTYLF 188
QY 189 GIGANISLLASGRAIWAASKLRASVSEGVDPWGISLRLLAMNTVSGLV---LMASYW 243
DB 189 SLFNSSLILIVGLFMMLNLSSEDTLIKFKFISDSKITLVGVSTPIIAIYALICCLVRF- 247
QY 244 WINKSVLTDPRFVNPSEMOKGKAKRKNMKOSFLYLDSPYLLITLLIVAGYCINL 303
DB 248 -ISKRYITNPLFYH--KTYSSRSSTAQRMGLIKSFYIKSKYLMILLICSAFGAINTL 303
QY 304 IEVWAKQOLKIQVPMNDYSEPMGNFSGFWGVSVLIMLFEVGNVIRKFGMLGALVTPV 363
DB 304 VEAWKAKIKELIPTYVTVAEFNSLYILWVG--VAIIMTIGNNVRMHNMFVAAVISPV 362
QY 364 WVLITGVIFPALVIFRNAQ---SGLVAMFGTTPMLAVVGAIONILSKSTYALPDST 419
DB 363 IIMVTGVLFGELIVFDQIILSLFDGAILM--SPLALAVSIGIGIONILAKGTXYSDWIS 419
QY 420 KEMAYIFLDDQOKKKGKRAALDVVAARFGSGGALIQGL--LVICSGIMPTYLAIVILL 477
DB 420 REMLYIFLDELDELTKTGGAAVDVISAKVGSSSGELVOSIIFTLVPAATFTSISPLWVFT 479
QY 478 FIIMPLVVSATKL 490
DB 480 FVCFAMPIYAVRKI 492

RESULT 21
B97765
ADP carrier protein (imported) - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #ext_change 22-Oct-2001
C:Accession: B97765
R:Ogata, H.; Audif, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; M01D:21442074; PMID:11557893
A:Accession: B97765
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-507 (XLR)
A:CROSS-References: GB:AE006914; PIDN:AA03060.1; PID:g15619599; GSPDB:GN00173
C:Genetics:
A:Gene: tlc2
C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 31.4%; Score 827; DB 2; Length 507;
Best Local Similarity 35.8%; Pred. No. 1e-55;
Matches 179; Conservative 106; Mismatches 183; Indels 32; Gaps 9;

QY 11 KLSFLPMTHTHEIKVLPMFLMFCITENVYTLVDRDTKTLIVAGSGSGEALPIFKWL 70
DB 11 KFRHIVPINSYELTFIPALMLPILNQLNVRISIKOSFVYTLISS--EVLSTFKMG 75

```

QY 71 VPCAIIFMLIYAKSINLSKQALFYAAGTFFLIFFALFPVIAIPLRDVLPTEFADRLQ 130
Db 76 EMPGILFVFIYSKLCINIMTEQVFRITITGFLEFFALFGILFPYAFHFHDEBLKH 135
QY 131 AILPGLGLVAILRNNTFAAFYVLAELIWSVMLSMFEGFANEITKIHEAKRFYALFGI 190
Db 136 ITVPLHKWFIINGOWSLVIFYIMGELMPVIFETLLYWQLANKITKEEAPREYSEFTL 195
QY 191 GANISLASGAIYMAK-----LRASVSEGDPNGISLRLLMANTIVSGVLMAASYWIN 246
Db 196 FQOTNLISGVIIIFYPAKSEHFLPLPFSHLNDTNEILKSFITYILLISGLCLALHKID 255
QY 247 KNVLTDBRFYVBEEMOKGKGAKP-----KNNKDSPLYLDRSPYILLTLTLVAYGICI 301
Db 256 KSVV-----EADKNIKFKQNRMDILKSLVDSAKVILTSXYLIGTICLWMSYMSI 306
QY 302 NLIEVYWSQKLQYPMANDYSEFMGNFSFMTGVSVILMLFVGANVIRKGMLTGALVT 361
Db 307 SLIEGLMWSKQKQOYPATKDFIAVHGKVFMTGILT--LVSAFLSSSLIRICGFWGAIIT 365
QY 362 PVMVLITGVIFPALVIFPNQASGLVAMPG--TTPMLAVVAGAIONILSKSTRYALPSTK 420
Db 366 PIMHFGAGVMEFSTFVPEHNLCINVTNLGYSAPLVIIVIFIGLMAHVLSSKYSKYSLPDATK 425
QY 421 EMAYIPLDQOKVKGKAAIDVVAARFGKSGALIQQGLVITCGSI--GAMTPYLAIVLLF 478
Db 426 EMVYIPLDSEKTKGKAADVWGAKIGKSGIAITQ---FISFIFPAVAINHNDIAGLIME 481
QY 479 ----IAILVLSAKINKLFF 494
Db 482 SFIIYCLMLGYGVLSKYY 501

```

```

RESULT 22
B71695
add. ATP carrier protein (t1c2) RP377 - Rickettsia prowazekii
C/Specials: Rickettsia prowazekii
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C/Accession: B71695
R/Anderson, S.G.E.; Zomorodipour, A.; Anderson, J.O.; Sichevitz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A/Reference number: A71650; MUID:99039499; PMID:982893
A/Accession: B71695
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-507 <AND>
A/Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14836.1; PID:g3860923
A/Experimental source: strain Madrid E
C/Genetics:
A/Gene: t1c2; RP377
;Superfamily: rickettsial-type ATP/ADP translocase

```

[illegible]

Qy	244	WINKVLDP	PRYNEEMOKKKGAKP	-----KNNKDSFLYDRSPYLLTLTLVAYG	298
		: : :	: : :	: : :	
Db	253	LIDKSV	-----EADKNIKFKNQRTDILKLSLESAKILITSYLDGFCILLVMSYS	303	
Qy	299	ICINLIEV	TSKSQLQYPRNNDYSEFMGNFSFWGVSVLMLFVGNNVIRKFGMLTGA	358	
		: : :	: : :	: : :	
Db	304	MSINLIEG	MSKXVQOLYPATDPLSYHGEVFWTIVLT-LVSAFLSSLRIRIGWEMGA	362	
		: : :	: : :	: : :	
Qy	359	LVTPLMVL	LTGVPFALVYFRNQASGLVAMFG-TTEPLMLAVVYGAIINITSKRYALFD	417	
		: : :	: : :	: : :	
Db	363	IITPLMEV	VAGMFSPTLFEQHLGNIVNTLGYSSPLVIVF1GGLMHWPAKSKYKLSFD	422	
		: : :	: : :	: : :	
Qy	418	STKEMAV	PLDOEKVKGAALDVAAFPKSGGALIQ	-----OGLVYC	462
		: : :	: : :	: : :	
Db	423	ATKEKENV	PLDNEIKTKGAADVGMARKIGSIGALIOFISFSIFPNAVHNDIAGLT---	479	
		: : :	: : :	: : :	
Qy	463	GSIGAMT	PLAVITLPLTIAIWLVSATKRL	493	
		: : :	: : :	: : :	
Db	480	-----MUTFIIV	CILMLYGVKVL--QNTYKMM	504	

RESULT 23
H82642
hypothetical protein XF1738 [imported] - *Xylella fastidiosa* (strain 9asc)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: H82642
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: H82642
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <SIM>
A:Cross-references: GB:AE003997; GB:AE003849; NID:G9106805; PIDN:AF84547.1; GSPDB:GN001.
A:Experimental source: strain 9asc
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brites, W.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krüger, J.E.; Kuramae, E.E.; Laiz
Chado, M.A.; Madeira, E.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: XF1738

[illegible]

Db 149 FDSEGRKRLFGFAAGAT----AGS---LGSATISGLIHLDREW-----LMAIAVIF 195

Qy 234 SGLVIMASYMWINNVLTDPFY-----NPEMGKKKAKPKKMKDSFLYIDRSFYI 287

Db 196 LELAVIAS-----RLSRIPAFQHAARDNPDPGLG-----GIPAGWHILRSPLYL 243

Qy 288 LLTLTLVAVAGICINLIEVTKWSQKLOYPNMNDSEFMGNFSFMTGVSVLIMLFGVGN 347

Db 244 GGLAIFILLYSVTSTFLYFQOASIAQASFPPDAATAPAFANIDLIVNAVTLVFPQLFVTRG 303

Qy 348 VIRKGMVLGALVTPFVMTLTGIVFPFALVIFRNQASGLVAMFGTPTLMLAVVGAIONIL 407

Db 304 MMAATGVIAATLCVLPVLSL-----GFAAL-ASP-SVAIVIVAA--QVA 343

Qy 408 SKSTKYALFDSTKEMAYIPLDQOKVKGAIDVVAARFGSGGALLIOGGLLVICGSIQA 467

Db 344 RRVANFALARPARELFTSSAREDDYKAKNFIDVYVGGQIQIISWGYAGLMGLT-GLTLA 402

Qy 468 MTPYIAVLLFTIIMVLSATKNTKLFIAQSALKKEQVADSDAPASS 515

Db 403 QIPMLAVPL---SAVWLGLSVWLGRTHQAQE-----RQDAATAPS 439

RESULT 24

hypothetical protein CT234 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jun-2000
 C/Accession: F71541
 R/Stephen, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A/Reference number: A71570; MUID:9900809; PMID:9784136
 A/Accession: F71541
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-928 <ARN>
 A/Cross-References: GB:AE001296; GB:AE001273; NID:93328630; PIDN:AA67826.1; PID:9332864
 A/Experimental source: serotype D, strain UW-3/Cx
 C/Genetics:
 A/Gene: CT234
 C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 6.6%; Score 173.5; DB 2; Length 928;
 Best Local Similarity 19.9%; Pred. No. 2.2e-05;
 Matches 106; Conservative 88; Mismatches 195; Indels 143; Gaps 22;

Qy 7 KPFGKLRG--FLMPHTHELKVKVLPMPFPCIFENYTVLRDTKOTLIVGAPG----- 58

Db 6 KGGSLRLRLRFLLSIRGEEKRAL-LFLLLGLI-----WSVACYGSIALGES 51

Qy 59 -----GAELAPF-----IKFWLVVPCAIIFMLIVAKSLNISKOALFYAVGTPFLI----- 104

Db 52 LFLBEIGAEKLPFAVLGASFPL---CFISCLILVNLRSKRVSPKALFLSFISCVLINCILY 108

Qy 105 -FPALFPVIVYPLRDVHLPTERAPRLQALIRPGLIGLVAIIRMTFAAFYTLAEIMG-SV 162

Db 109 LFMHL-----AIH-----KGVSGPTFL-----YRLI--IMGLTI 136

Qy 163 MLSLMFMGFANETIKHAKRFPVALFGIANISLASGAIWMAK-LRAVSEGVDPWG 221

Db 137 LCYANFMGFIQDFNIDQAKHFCIF-----NALTFCGDFGAIIVNOIQYILG 184

Qy 222 ISLRLAMMTIVSGVLVMAASYMWINNVLTDPF-----FYNPEMGKKKAKPKKMMK 275

Db 185 AELIILARIYITFIFPLVHTYISSLSKELSEHDHFLDTGYRPSRKQ-----TLKICLK 238

Qy 276 DSFLYLDSPYILTLTLVAVAGICINLIEVTKWSQKLOYPNMN--DYSEFMGNFSFWT 333

Db 239 DKYTFYLVSYFYLQGLLVFT-----ERNYIKIFDAQGNETCELTENFTKYSWI 290

Qy 334 GVAVVLIMLFGVGNVIRKEGMLTGALVTPVNVLLTGIVFPALVIFRNQASGLVAMFGTTP 393

Db 291 SLGNMFPALPAYSVRITRFG-----INNIIILFAPICF-----PSLFCQWS 330

Qy 394 LMLAVVGAIONILSKSTKYALFDSTKEMAYIPLDQOKVKGAIDVVAARFGSGGAL 453

Db 331 IKTSVFATMGMIAREGLAVALDDNNLQILYIGPKNIRNQVRAIE-----SF 379

Qy 454 IQGGLVICSIGMTPYLAIVIIILFIAMLVSA-----TKLNTKLFIAQSALK 501

Db 380 VEPAGMFIALLCLFIPHOYVLCIISAVCILALIRTHSKAILNLSIE 431

RESULT 25

conserved hypothetical protein TC0505 (imported) - Chlamydia muridarum (strain N:9g)
 C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
 C/Accession: B81695
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, F.C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: B81695
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-918 <TET>
 A/Cross-References: GB:AE002319; GB:AE002160; NID:97190541; PIDN:AAF39347.1; PID:9719054
 A/Experimental source: strain N19g (MoPn)
 C/Genetics:
 A/Gene: TC0505
 C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 6.1%; Score 161; DB 2; Length 918;
 Best Local Similarity 19.0%; Pred. No. 0.0002;
 Matches 88; Conservative 80; Mismatches 198; Indels 96; Gaps 14;

Qy 57 GSGEALPFIKPMVLPVPCAIIFMLIVAK---LSNIJSKOALFYAVGTPFLIFALPFTVI 113

Db 13 GEEKRALPFLILGLWSVACYSIALGESLPLEBIGTKLPFAVLGASFPLCF-ISCVL 71

Qy 114 YPLRDVLPTEFADRLQALIRPGLIGLVAIIRMTFAAFVLA-----EL 158

Db 72 YSL-----SRKASPKALFLSFISCVLNCNLVLPFHLAIHRGVSGPTFLVRIIL 121

Qy 159 WG-SVMLSMFMGFANETIKHAKRFPVALFGIANISLASGAIWMAK-LRASVSEG 216

Db 122 WGLTILCYANFMGFWVDQFNIIDAKHFCIF-----NALTFCGDFGARIWQ 169

Qy 217 VDPWGISLRLAMMTIVSGVLVMAASYMWINNVLTDPF-----FYNPEMGKKKAKP 270

Db 170 IQHLGAEILILSFIAVITFPLVHTYISSLSKELSEHDHFLDTGYRPSAKQAF----- 223

Qy 271 KMMNKDSFLYLDSPYILTLTLVAVAGICINLIEVTKWSQKLOYPNMNDY--SEFMGN 328

Db 224 KLCKDKYTFYLVSYFYLQGLLVFT-----ERNYIKIFDAQFNAKNFELTENFTK 275

Qy 329 FSEFTGVASVILMLFGVGNVIRKEGMLTGALVTPVNVLLTGIVFPALVIFRNQASGLVAM 388

Db 276 YSSWISLGNMFPALPAYSVRITRFG-----INNIIILFAPICF-----FSL 315

Qy 389 FGTPPLMLAVVGAIONILSKSTKYALFDSTKEMAYIPLDQOKVKGAIDVVAARFGK 448

Db 316 FVCKSITSVLVINTMGMIAREGLAVALDDNNLQILYIGPKNIRNQVRAIE----- 367

Qy 449 SGGALLIOGGLVICSIGMTPYLAIVIIILFIAMLVSAATKL 490

Db 368 ---SFVEPAGMFIALLCLFIPHOYVLCIISAVCILALIRTHSKAILNLSIE 406

RESULT 26

A72096
 ct234 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C|Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C|Accession: A72096
R|Kaitman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A|Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A|Reference number: A72000; MUID:99206606; PMID:10192388
A|Accession: A72096
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-925 <ARN>
A|Cross-references: GB:AE001614; GB:AE001363; NID:G4376562; PIDN:AAD18442.1; PID:G437656
C|Genetics:
A|Experimental source: strain CWL029
C|Gene: CPN0293
C|Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 6.0%; Score 158.5; DB 2; Length 925;
Best Local Similarity 20.6%; Pred. No. 0.00331;
Matches 95; Conservative 94; Mismatches 165; Indels 109; Gaps 26;

OY 59 GAELPFIKFWL---VPCAIIFMLIYAKLSNIISKQALF-YAVGTPLIFPALFPVIY 114
DB 49 GSAELP-KIYLGSLLICVLSSLILYNMFKHISATAPLPISLTLCNFYILTSEIF 106
OY 115 PLRDVHTFEFRADRLQALLPPGLGLVALINNTWPAAYVALELMGSYMLS-LMFNGPAN 173
DB 107 -----AIDPP-----RSPLF--FYRIIV-TWSLTILSYTSFGWFVD 138
OY 174 EITKHEAKRFYALEFGIGANISILASGRAIWASCLRASVEGYDPWGDISRLMAATIY 233
DB 139 QFNLDQDKRHFCIN-----AIIPLGDPI-GSSITSLVHTIGIQILLIPNALVLT 191
OY 234 SGVLIMASYWMINKNV--LTDPR--FYNPENQKGKKGAKPKMMKOSFLYLDSPYIL 288
DB 192 FPIV----FYVSKSLKSLSDHDHDFID-----TGHPPPLSKALKCFYDKTYFTLL 236
OY 289 ----LTLTLVAYGCINLIEVTMSQAKLOYPNNNDYS--EEMGNBSFMGVSVYLIML 342
DB 239 CFYFMQLLAIA-----TEFNLIKIFEIQPASKEEFELVAHI GKCSLWISLGNCFFAL 291
OY 343 FVGGNVIRKPFMLTGALTVPWVVLLTGIVGFALVFIRQAAGLVAMFGTPTPLMALVAVGA 402
DB 292 PAYSIRIVRLG-----VNMIILPAPLCFLS-----LFLPTFTTTLSIAVLA 333
OY 403 IONILSKSTKYALPDSTEKMAV--IPLDQEKVKYKAADIVVAARFGSGGALLIQGGILV 460
DB 334 M--VVRGEVTVALDDNNQLLIYGVP---NKIRNQ--IRIVPSFTPIGLMWLS---L 382
OY 461 ICGSIGMTPLIATVILFTIALMWLSATKANLKPLFAQSALK 501
DB 383 IC--FLSSOOYVFCIIISLIAITIVC---LVRSYYAAKAILK 418

RESULT 27
EB1573
conserved hypothetical protein CP0465 [imported] - Chlamydophila pneumoniae (strain ARJ33)
C|Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C|Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
C|Accession: EB1573
R|Red, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, K.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A|Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A|Reference number: AB1500; MUID:20150255; PMID:10664935
A|Accession: EB1573
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-925 <REN>
A|Cross-references: GB:AE002208; GB:AE002161; NID:G7189387; PIDN:AAF38302.1; PID:G718938
A|Experimental source: strain AR39, HL cells
C|Genetics:
A|Gene: CP0465
C|Superfamily: Chlamydia trachomatis hypothetical protein CT234

[illegible]

```

Qy 234 SGVLVMSYMWINKV--LTPDR--FYNPEMQKKKAKPKMMKDSFLYLDSPYL- 288
Db 192 FPIV-----FVYSKSLKSDHDFID-----THPPPLSAALKCYDKTFLL 238
Qy 289 ----LTLVLVAYGICINLIEVTKSQKLOYPNNNDYS--EFMGNFSFMTGVSVLIML 342
Db 239 CFYFLMQLALTA-----TEFNLIKIFEIQPASEEELVAHIKCSLMISLGMCPAL 291
Qy 343 FVGGNVIRKFGMLTGALVTPVWVLTIGIVFPALVIFRNQASGLVAMFGTTPMLAVVGA 402
Db 292 FAYSRIVARLG-----VNNIILFAPLCFLS-----LFLFWTFKTLISIAVLA 333
Qy 403 IONILSKTKVALPDSTKEMAY--IPLDQEKVKKKAIDVVAARFGSGGALLQGGILV 460
Db 334 M--VRESEVVALDNNIQLILYGP---NKIRNQ--IRLVESFIEPIGLMWS--L 382
Qy 461 ICGSIGAMTPYLAIVLFLIAIMIVSATKLNKLFLAQSALK 501
Db 383 VC--FLSSQYVFCILISLILVLC--LVRSYAKAKILK 418

```

RESULT 29

```

AF2383
hypochemical protein al14622 [imported] - Noctoc sp. (strain PCC 7120)
C/Species: Noctoc sp. PCC 7120
A/Note: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AF2383
R/Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Matanabe, A.; Iriyuchi,
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A/Reference number: AB1807; MUID:21595285; EMBL:U1759840
A/Accession: AF2383
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1002 <KIR>
A/Cross-references: GB:BA000019; PIDN:BA076321.1; PID:G17133759; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
C/Gene: al14622
C/Superfamily: Chlamydia trachomatis hypochemical protein CT234

```

Query Match 5.7%; Score 151; DB 2; Length 1002;

Best Local Similarity 20.2%; Pred. No. 0.0013;

Matches 101; Conservative 88; Mismatches 196; Indels 114; Gaps 23;

```

Qy 19 IHTHELKKVLPMPFMFCITFNVTYVLRDXTLLIVGARGSGAEALPFIKFWLVPCALIF 78
Db 24 LRPEEGERTMMFAFYTTVSIGLRKADSTVALFLDEYGAG--PLP---MMYIASAVMG 77
Qy 79 M---LIYAKLSNIISKQALFYAVGTPFLIFPALPFTVYLPRLDVLHPTFADRLQALIP 135
Db 78 MALVFVYSWMLQKIFPLR-----WVVAIAPCMIVPL-----ILLVLLRW 116
Qy 136 GL-----LGLVALLRNTTFAAFYVLAELNGSVMLSLMFGPANEITIKHAKRFPALFG 189
Db 117 GIDVAVFSVIVFLRLVW--DSIYVNDLNTSIV-----ANQLFIRIKRTYPLIS 167
Qy 190 IGANISILASGRATVMA---SKLRASVSEGVDPWGISLRLM---AMTYISGLVMSY 242
Db 168 SGLLVADVISGFSPLMLLEPAKN-----RVIMACGVMTISAILCHYST 213
Qy 243 WINKNVLTDRFYNPEMQKKKAKPKMMKDSFLYLDSPYLTLTLVLVAYGICIN 302
Db 214 QYRTSPFPAQRLLPOEQASHRRRIQAP-----LKR--YTLQFAVVGILQIIGL 261
Qy 303 LIETWKSQKLOYPNNND--YSEPMGNFSFMTGVSVLIMLFGVGNVIRKFGMLTGALV 360
Db 262 LVDPQYLOELKI---NIGDRELAGFLGIGIVGICELTTQMFVSSRLIERFGVFTPAL 318
Qy 361 TPVWVLLTGIYFPALVIFRNQASGLVAMFGTTPMLAVVGA--QNILSKTKVALPDST 419

```

```

Db 319 LPVAV---GGVYPEMIVVLLBLEIGS-----LAFPMGLVKRFEDLRTFVISS 367
Qy 420 KENAYITLDOEQYKKAIDVVAARFGSGGALLQGGILVCSISGAMTPYLAIVLFI 479
Db 368 GPLYQPI--PERISRTOV-----LSGGT---AEAIAGTAG-----IIVITLFFV 409
Qy 480 IAIWLSATKLNKLFLAQS 498
Db 410 CGLF-VEPATMOKWFISET 427

```

RESULT 30

```

B82219
transporter, BCTT family VCI279 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: B82219
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; EMBL:U0952301
A/Accession: B82219
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-540 <HEI>
A/Cross-references: GB:AE004207; GB:AE003852; NID:99655761; PIDN:AAF94438.1; GSPDB:GN001;
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VCI279
A/Map position: 1
C/Superfamily: Escherichia coli probable carnitine transport protein

```

Query Match 5.5%; Score 145; DB 2; Length 540;

Best Local Similarity 20.5%; Pred. No. 0.0018;

Matches 106; Conservative 76; Mismatches 187; Indels 148; Gaps 24;

```

Qy 69 WLVPVCAIIMLI-----YAKL-----SNLSKQALFYAVGTRP--LIFPALF 109
Db 79 WLFVWSGNLFLFLCVGLVISPFGIRGQAKADHSFLSLMMLFPAAGGIGLGMFWSVA 138
Qy 110 PTYVY---PLRDVLHPTFADRLQALPLGL--LGLVAILRNTTFAAFYVLAELMGS 161
Db 139 EPVAYITGMYETPL-----GVEANTPQAQLAMGATVHFHGLHW---AIYGV 183
Qy 162 VMLSLMF-----WGFANEITIKHAKRFPALFGISANISL-- 196
Db 184 VALSLAFPCYKNGKGLPLSMRSVFYPLDGRAMGNAGHVLDLAV--LATLFGLATSLGLGA 241
Qy 197 --LASGRAIYASKLKASVSEGVDPWGISLRLMANTYISGLVLMASYMWINKNVLTDR 254
Db 242 QQAASGIHHVFGFES-----GIGLOIAY--ICVVGTLATLSVVRGID----- 281
Qy 255 FYNPEMQKKKAKPKMMKDSFLYLDSPYLTLTLVLVAYGICINLIEVTKSQKLO 314
Db 282 -----GGVKVINSINMIVAFI-----LALVAIVIGSVITLSIPTLIMAYLQN 324
Qy 315 QYPMNDYSE-----FMGNFSF-----WTGVSVLIMLFGVGNVIRKFGMLTGALVTP 362
Db 325 LIPLSNHRGREDSEWFGGVVFWYAMWISSPVGMFIYAVSGRTIRF--IYAVLLIP 382
Qy 363 VMQLLGTI--VFPALVIFR--NQASGLVAMGTTPLMLA-----VVGAIQNILSKSTK 412
Db 383 TLVTLWMSVFGGLAVDQVINEI--GVLGGNGLTDVSLAMPQMEDSLVFGVLSIAVAVL 441
Qy 413 YALPDSKEMAYVPLDQ-----EOKYVSKAIDVVAARFGSGGALLQGGILVI 461
Db 442 LVFFITSSDGLVDSITSGKLDSEVQLQKVWAVLEGAAVVALMTGGTEAVQALQ-- 499
Qy 462 CGSIGAMTPYLAIVLFLIAIMIVSATKLNKLFLAQS 498

```


Db 500 AGVISTALPFTVILLMVCVSLM--KGLKTEPEFMAS 534

RESULT 31

Probable transmembrane transport protein NMA1483 [imported] - Neisseria meningitidis (str E81839)
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: E81839
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, H.; Holtz, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: A81775; MUID:20222556; PMID:10761919
 A:Accession: E81839
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-660 <PAR>
 A:Cross-references: GB:AL162756; GB:AL157959; NID:G7380091; PID:CA884716.1; PID:G738013
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1483
 C:Superfamily: choline transport protein betr

Query Match 5.5%; Score 144; DB 2; Length 660;
 Best Local Similarity 20.7%; Pred. No. 0.0027;
 Matches 127; Conservative 99; Mismatches 171; Indels 216; Gaps 34;

```

OY 32 LMFCTITNTVLTADTDTLVGAPG-----SGAEAIPIKF-WLVVPCAIIFM----- 79
DB 7 LFFVCVIV-----VLVLTVPDQVOMWLDRAKEVIFTEFSWFVLTFSIFIGFLII 56
OY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALPPTVYPLRDVHPTEF 125
DB 57 LVSLSLGNIRIGRDEVDPEFGFLSWLMLFPAAGVGVLMFGVAEPLMHTYSDITAGTPE 116
OY 126 ADRLQATILPGLIGLVAILRNWTFAPFYVLAELMGVSLIMFGFANEITKIHAKR-- 183
DB 117 HROQOALL-----HTVFHMGVHANSV-----YGTIALALAYGFR--YKLPALRSC 161
OY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKRASVSE-GVDPWGISRL 227
DB 162 FYPLTKERISGRFGDAIDIMLATPFGIITTLGFGASQLGAGLQEIWIENSFSVQVL 221
OY 228 MAMTIVSGVLMASVWYINKNVLTDRPFYNPEEMQKKGAK--PKNMKDSFL-----YL 281
DB 222 IIAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLFLFVL 261
OY 282 DRSPYIILLTLLVIAVGICI-----NLIEVTWKSQKLQYPMNDYSBFGNFSF----- 331
DB 262 AAGFTVILS-----AFGDNIGNYIGNLVRLSFKT-----YAYERHKEPWFESWTVLYWAM 312
OY 332 ---WTGVSVLIMLFVGVGNIRKFGMLTGAIVTPVWMLLTGIVFPALVIFRNOASGLVAM 388
DB 313 WGSNAPFVGFLIARISKRTIRER--VFGVLLIPG--LFGVLMF-----TV 354
OY 389 FGTTPMLV-AVYVGAIONILSKSTKVALPDSTKEMAYIPLDQEKV----- 433
DB 355 FGNATIMLNDGVAGVLEKMTSSPETLLF--KEFNLYLPBELTISVLSLVISLFPYVTA 411
OY 434 -----KGAIA-----IDVVAARFGSGGALIQGGLVIGSIGAM 468
DB 412 DSGIYVANNITSRKGLSAPRQAVMGMVLSAVALMLMRSGG-----LGNLSM 461
OY 469 T-----PYLAVILFIITAIWL-VSA-----TKLN-----KLFASALKEQ 503
DB 462 TLIVSLPFLALMLIMCFSLMKGLSADKCYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 521
OY 504 EV-----AQEDSAPA 513
DB 522 DILKFLKHTASPA 534

```

RESULT 32

transporter, BCC1 family NMB1277 [imported] - Neisseria meningitidis (strain MC58 serogr E81101)
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: E81101
 R:Jettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A. Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; et al.; Olt, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: E81101
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-675 <TET>
 A:Cross-references: GB:AE002476; GB:AE002098; NID:G7226514; PID:AAF1653.1; PID:G722651
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB1277
 C:Superfamily: choline transport protein betr

Query Match 5.5%; Score 144; DB 2; Length 675;
 Best Local Similarity 20.7%; Pred. No. 0.0028;
 Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;

```

OY 32 LMFCTITNTVLTADTDTLVGAPG-----SGAEAIPIKF-WLVVPCAIIFM----- 79
DB 22 LFFVCVIV-----VLVLTVPDQVOMWLDRAKEVIFTEFSWFVLTFSIFIGFLII 71
OY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALPPTVYPLRDVHPTEF 125
DB 72 LVSLSLGNIRIGRDEVDPEFGFLSWLMLFPAAGVGVLMFGVAEPLMHTYSDITAGTPE 131
OY 126 ADRLQATILPGLIGLVAILRNWTFAPFYVLAELMGVSLIMFGFANEITKIHAKR-- 183
DB 132 HROQOALL-----HTVFHMGVHANSV-----YGTIALALAYGFR--YKLPALRSC 176
OY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKRASVSE-GVDPWGISRL 227
DB 177 FYPLTKERISGRFGDAIDIMLATPFGIITTLGFGASQLGAGLQEIWIENSFSVQVL 236
OY 228 MAMTIVSGVLMASVWYINKNVLTDRPFYNPEEMQKKGAK--PKNMKDSFL-----YL 281
DB 237 IIAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLFLFVL 276
OY 282 DRSPYIILLTLLVIAVGICI-----NLIEVTWKSQKLQYPMNDYSBFGNFSF----- 331
DB 277 AAGFTVILS-----AFGDNIGNYIGNLVRLSFKT-----YAYERHKEPWFESWTVLYWAM 327
OY 332 ---WTGVSVLIMLFVGVGNIRKFGMLTGAIVTPVWMLLTGIVFPALVIFRNOASGLVAM 388
DB 328 WGSNAPFVGFLIARISKRTIRER--VFGVLLIPG--LFGVLMF-----TV 369
OY 389 FGTTPMLV-AVYVGAIONILSKSTKVALPDSTKEMAYIPLDQEKV----- 433
DB 370 FGNATIMLNDGVAGVLEKMTSSPETLLF--KEFNLYLPBELTISVLSLVISLFPYVTA 426
OY 434 -----KGAIA-----IDVVAARFGSGGALIQGGLVIGSIGAM 468
DB 427 DSGIYVANNITSRKGLSAPRQAVMGMVLSAVALMLMRSGG-----LGNLSM 476
OY 469 T-----PYLAVILFIITAIWL-VSA-----TKLN-----KLFASALKEQ 503
DB 477 TLIVSLPFLALMLIMCFSLMKGLSADKCYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 536
OY 504 EV-----AQEDSAPA 513
DB 537 DILKFLKHTASPA 549

```

RESULT 33

B83719
multidrug resistance protein (efflux transporter) BHO54 [imported] - Bacillus halodurans
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 06-Jan-2003
C:Accession: B83719
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; M01D:20512582; PMID:11058132
A:Accession: B83719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AF001508; GB:BA000004; NID:g10172890; PIDN:BA804273.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BHO54
C:Superfamily: multidrug-efflux transporter

Query Match 5.3%; Score 140; DB 2; Length 474;
Best Local Similarity 22.7%; Pred. No. 0.0039;
Matches 110; Conservative 74; Mismatches 212; Indels 88; Gaps 22;

QY 67 KFWLVPCALIFMLIYAKLSNLSKQAL-----FYAVGTPPII-PPALPPTVYIPR 117
DB 6 KKMVVV-CVILFGSTMTILNMSMLNPAIFQLMNVBADAVATGWTTFPMVAMGTMPT 64
QY 118 DVLHPTERADRLQALIPGLGLVAIL--RMNTFAFYV---LAEIWSVM---LSIMF 168
DB 65 GYL-GDKIGKKEAYILGIFVLGSLGALSNLPSLIVFRLQIGGGVMMPLSMTLIF 123
QY 169 -----WGFANEITK-----IHAKEFPYALFGANISLLASRAIVW 205
DB 124 DAFPRNEKGLATGVGVASVMAPIGPTLGGFIVTSNMKYLELVNIPFGLIGIIAAVY 183
QY 206 ASKLRASVSEGVDPFGISL-----RLMAATVSGVLMSYMWINKVLTDPFRNPE 259
DB 184 LPKIARSGQIKIDRNGPLFVTAGVGSILAFGRMDLTHLT- WINGVLLILGLCLV 241
QY 260 EMQKKGKAKPRQMKNSFLYLDSPYLLLTLLVIAVGI-CINLIEVTKSQKLQYRN 318
DB 242 FVQVEKRAEQPLD-----LSLFRIP--AVSLISWVAGISSIGFAGIFVPLLLQ--Q 291
QY 319 MNDYSEPMGNBSFMTGVSVLIMLVGVGVIRKFGMLGALVTPMV-----L 366
DB 232 VYDVPIMTGLVFLPSALFTGLTMSIGRMIDKRG-PGIMTAGMMAVAGTFALGYLHL 350
QY 367 LTGIYF-FALVIFRQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDTKEMAYI 425
DB 351 ETGLMYITAFMAAIRGVGGLTTPATTGMAIIPGGLISRGSAAMNVLRQSSAGIYFI 410
QY 426 PLDQKQYKKAIDVVAARFGSGGALIQGLLVICSGIGMTPYLAIVLLFIATLV 485
DB 411 SVFFEVF-RGQAL--VSTSFPEATLQAINEGFF-----VGFLT-----ALSIAPAWLE 458
QY 486 SATK 489
DB 459 KKA 462

RESULT 34

T07946
cytochrome-c oxidase (EC 1.9.3.1) chain I - Cryptocodinium cohnii mitochondrion
C:Species: mitochondrion Cryptocodinium cohnii
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T07946
R:Norman, J.E.; Gray, M.W.
FEBS Lett. 413, 333-338, 1997
A:Title: The cytochrome oxidase subunit I gene (coxI) from the dinoflagellate, Cryptococ
A:Reference number: Z16237; M01D:97424386; PMID:9280308
A:Accession: T07946
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-536 <NOR>
A:Cross-references: EMBL:AF012554; NID:g2358288; PIDN:AA869658.1; PID:g2358289
A:Experimental source: strain WH-d
C:Genetics:
A:Gene: coxI
A:Genome: mitochondrion
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: cytochrome-c oxidase chain I homology <COI>
F:22-479/Domain: cytochrome-c oxidase chain I homology <COI>
F:75,395/Binding site: heme a iron (His) (axial ligands) #status predicted
F:255,304,305/Binding site: copper (His) #status predicted
F:255-259/Cross-link: 1'-histidyl-3'-cytosine (His-Tyr) #status predicted
F:259/Binding site: oxygen (Tyr) #status predicted
F:385/Binding site: magnesium (His) (shared with chain II) #status predicted

Query Match 5.2%; Score 137; DB 2; Length 536;
Best Local Similarity 20.3%; Pred. No. 0.0075;
Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;

QY 68 FWLVPCCALIFMLIYAKLSNLSKQALFYAVGTPPII-PPALPPTVYIPR 127
DB 81 FYLVMP-----GLFGGNGYFIVFQSPFVVP----- 109
QY 128 RLQ-AIIPGLGLVAIL-----RMVTF-----AAFYVA-ELIWSVMSLMEWG 170
DB 110 RVNPFSLILSLTLLILSLISFEGGTGTLXPPLSTPMNLSPSISGNIIFGLISG 169
QY 171 FANETIKIHEKRPYALFGIGANISLLASGRAIVWASKLRASVSEGVDPWGI---SLRL 227
DB 170 ISSVLTSLNFWITLSLSISGITLTKTS-----LFPWLLITSGMLL 211
QY 228 MAMTIVSGVLMSYMWINKVLTDPFRFYNPEMQKKGAKPRQMKNSFLYLDSPY 286
DB 212 LTLPLTGLALMT-LSDINVTLPFDFP-----GGDIFVQHLFWFPGHPEVY 259
QY 287 ILLP-----TLVIAVGI-----CINLI-EVTKSQKLQYRNNDYSEF 325
DB 260 ILIIPAGIISIIISGLQIKIIPGNPSMIFPMSISILSGVWGHNVYITGLEDTDSYF 319
QY 326 MGNFSFMTGVSVLIMLVGVGVIRKFGMLTGAIVTPVMTL-----GIVFALVIFR 379
DB 320 SG-----VTIISLPTGFKI---FNMVSTYIGNPSLLILTKNSSLFGLFL-LMFTI 367
QY 380 NQASGLVAMFGTTPMLAV-----VGAIONILSKSTKVALFDTKEMAYIPDQEQVK 434
DB 368 GGSYGII--IGNAAVDGLDHTYIIAHFHFVLSGAVIAIFSG-----IIFNEXIKIG 419
QY 435 GKAIIDVVAARFGSGGALIQGLLVICG-----SIQ 466
DB 420 SKNLPSCSSNSRYNLVLPFGLITLFGPMHFLGFMVPRRIIDPFDSFISNPLSSIG 479
QY 467 AMTPYLAIVLLFIATLVLSATKINKFLAOSA 499
DB 480 SGITLSFGFLFKENCWITSIAQCSRLFLGHVA 512

RESULT 35

T14236
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Synechococcus sp. (strain PCC 70
C:Species: Synechococcus sp.
A:Variety: strain PCC 7002
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 03-Jun-2002
C:Accession: T14236
R:Klughammer, B.; Sultemeyer, D.; Badger, M.R.; Price, G.D.
submitted to the EMBL Data Library, April 1997
A:Description: Involvement of ndhF3, ndhD3 and ORF427 genes in high affinity CO2 uptake
A:Reference number: Z17936
A:Accession: T14236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-498 <KLU>
A:Cross-references: EMBL:U97516; NID:g2232044; PID:g2232047; PIDN:AA862185.1
A:Experimental source: strain PCC 7002

C:Genetics:
A:Note: nhd3
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 5.2%; Score 136.5; DB 2; Length 498;
Best Local Similarity 20.6%; Pred. No. 0.0076;
Matches 92; Conservative 60; Mismatches 139; Indels 155; Gaps 20;

QY 82 YAKLSNIIISKALFAVAGT-----PFLIFPALFPTV-IYPLRDVILHPTFADRLQAILPP 135
DB 55 YTEPHNMIISIIIGLNINLGVDSLSPLIVINSILTLVAIYSIGESNHRKLYSLILLINS 114
QY 136 GLGLVALLRMTTFAPFYVLAELMGSVLMLMFQFANEITKIHAKRFYALFGANIS 195
DB 115 GITG-ALIAN-----NLLPLFLFY---EI---ELIPFILLAL----- 145
QY 196 LLAGRAIWMASKLRAVSEGVDPWGISLRILMAMTIVSGVLMAVW---WINKVLLTD 252
DB 146 -----W-----GGEKKGVASTKFLIYTAISGLVLAIFGIYWLQSSNFD 186
QY 253 PRFNPEEMQKKGKAKPKMMKDSFLYLDSPYLLITLIVAVGICINLIEV-TWKSQ 311
DB 187 FENLTLEMLBENTK-----VILLTILLIGSGIKIPVPLHTW--- 223
QY 312 LKLOYPMNDYSEFMGNFSFWTGVSVYLIMLFVGNVIRKFGWLTGALVTVMLLIGIV 371
DB 224 LPDAVVEANP-----AVTVL-----GGVPAKLG-----TYGLV 252
QY 372 FFALVIFR---NOASGLVAMFGTTPMLAVVGAIONILSKTYALPDSTKEMAYIPLD 428
DB 253 RFGIQLFPDWSVTSPALAVIGTVSMVGSILAALAQRLKMWAYS---SIGHGVIIVS 309
QY 429 Q-----EOKYKKAIDVVAARFGKSGGALLIQ 456
DB 310 TAACTELSLGAVAQMISHSLIALLFHVLGIIERKV-GTRDLDVNLGMAMPVGLPLTS 368
QY 457 GLIVICSGIGMTF---YLAVILLF 478
DB 369 SLTLIGMASAGITGVLGVFAEFLVF 394

RESULT 36

C83615
Probable MRS transporter PA0241 (imported) - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83615
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bt
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83615
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <STO>
A:Cross-references: GB:AE004462; GB:AE004091; NID:g9946077; PIDN:AG03630.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0241

Query Match 5.1%; Score 135; DB 2; Length 441;
Best Local Similarity 21.1%; Pred. No. 0.0086;
Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26;

QY 32 IMFPCIFNTVAVDVTQTLIVG-----AGSGAEAPF--IKFVLVPCALIFMLIYA 83
DB 39 LVFAFTLNTYI-----DRALGVMOPIIAEKMSMTAMDYANINFWQVGVYAIFITLQGR 92
QY 84 KLSNLSKQALFYAVAGTFPLIFPALFPTVIVPLRDVILHPTFADRLQAILPPGLGLVAVI 143

DB 93 FIDKGVKRAFLVAV-----LLMSLATGAHGLATS 122
QY 144 LRMTFAAFYV-LAEL--WGSVMSLMPFGFANEITKIHAKRYA--LGICANISILA 198
DB 123 AAGFMVGRFIIIGLEAANYAPACVKTRLMPEAGE-----RAVAAGIFAGTNVGMV 174
QY 199 SGRAI-----VMASKLAASVSEGVDPWGISLRILMAMTIVSGVLMAVWINKVLLTD 253
DB 175 TPALLPLILGVW-----WQAAFICSAI--GLWVLV-FWVRN----- 209
QY 254 RFYNPEEMQKKGK-----AKPKMMKDSFLYLDSPYLLITLIVIA----- 296
DB 210 -YNNPEEHPKXQSELEFYIQQSEEPATRVFSGILRRGWAPALAVSTIAPVFWFYLY 268
QY 297 -----YGCINLIEVTKWSQLKQYPMNDYSEFMGNFSFWTGVSVYLIML----- 342
DB 269 WLPEFLNQYGLGISTVQM-----GIPLLIIMLTDAPG 301
QY 343 FVGGNVIRKFGWLTGALVTPVMTLIGIVFPAL-----VIFRNQSGI-VAMFGTTPMLA 397
DB 302 SVGGGILS--SWLIGRGPAPARLMSLLPACTIVGVFPAANSGLMIAV-----LAIA 354
QY 398 VVVGAIQ-----NLSKSTKYALPDSTKEMAYIPLDOEKYKKAIDVVAARFG---KS 449
DB 355 LAVGAHQAMTANIS-----LYMDYTPHGLMSTVFG-----FGMCAL 393
QY 450 GGALLIQGLVITCGSIGMTPLYAVIILFTIATWVSATKL 490
DB 394 GGMFTQ---IVGGVLTATNNNAVILFTMTIPAMVFIALTWL 431

RESULT 37

E75212
Oligosaccharyl transferase PAB2202 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: E75212
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: E75212
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-976 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CB49148.1; PID:9545765
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2202

Query Match 5.1%; Score 134; DB 2; Length 976;
Best Local Similarity 20.9%; Pred. No. 0.025;
Matches 107; Conservative 79; Mismatches 157; Indels 168; Gaps 29;

QY 59 GAELIPRIRKRWVAVPCALIMLIYA-----KLSNLSKQAL--F 95
DB 19 GKSLLYPLK-RILPLAVIGFIVAYYLRHLTAGKYFPDPTGYHFIYLVLEGLPKY 77
QY 96 Y-----AVGTPFLIFPALFPTVIVPLRDVILHPTFADRLQAIL--PP--GLGLVIA 142
DB 78 YPMAEAPFGSLIGEPGLY--ILPAITYKVSIVGNHF-----QALPMWPPFVGLGVIA 131
QY 143 -----ILRNWTFAPFYVLAELMGSVLMLMFQFANEITKIHAKRFYA--LFGIGA 192
DB 132 VYLLGRVYLVNEM-----AGLMAVILSVSTANFERTPSGNARCGDPMTLFLFSLVA 183
QY 193 NISLASSG--RAIVAS--KLRAVSEGV---DPWGISLRILMAMTIVSGVLMAVW 244
DB 184 MLVYLKENDIKKSLMWAVVLASISLGMWNGSPFGLMV-----LIGFASFQT 232
QY 245 INKNVLTDPFRFYNPEEMQKKGKAKPKMMKDSFLYLDSPYLLITLIVAVGICINLI 304
DB 233 IALFI-----FGKIKELKRVKFRYP-----AYALAIAGIGLTI--- 267

```

QY 305 EWTWKSQKLQGYPMNNDYSEPMGNFSF--MTGVSVLIMLFGVGNVIRKFGMLTALVTP 362
DB 268 -----PGIAKIGGFI-KPAFEVFLGLVLLVITMLYGGKFLVSDKKRFAVA 314
QY 363 VVLLTGVLFALVIFRNOASGLVMPFGTTPMLAVVGAIONILSKSTKVALPDSTKEM 422
DB 315 VIVILG---FA-----GAVAYVG--PKLFRMLMGAYO-----STQ--VVOGVDEL 352
QY 423 AVIPLDQKQKGAIDVVAARFGSGS-----ALIQGLLVTC-----SIGMT 469
DB 353 A-----KTLISDIKLYGVGEGNGLVFFLISIPFLIILGYLMALLKKSSSSN 400
QY 470 PYLAVILLFIATIMVNSATKMLKFLAQSAL 500
DB 401 EYMLSVFYIMSLYLJSLA-VAFPLASTYAI 430

RESULT 38
D64137
best protein homolog - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Sep-1999
C/Accession: D64137
R/Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
D.M.; Brandon, R.C.; Fine, L.D.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: D64137
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-669 <TIG>
A/Cross-references: GB:U32843; GB:U42023; NID:91574554; PIDN:AAC3352.1; PID:91574560; T
A/Experimental source: choline transport protein best
C/Superfamily: choline transport protein best

Query Match 5.1%; Score 133; DB 2; Length 669;
Best Local Similarity 19.8%; Pred. No. 0.019;
Matches 116; Conservative 87; Mismatches 167; Indels 216; Gaps 30;

QY 15 FLMPHTHEL-----KVLPMFLMFFCTFNTVTRDTKTLIVAGPSSGAALPPIKFW 69
DB 27 FIAPBQTALNQAKSGIFANFSWTVLTSV-----FLGFL 63
QY 70 LVVPCAILFMILYAKLSNI-----LSKQALFYAVGTFP-LIFPALFTVITP 115
DB 64 LI-----LSVSSIGNIKLGQDEEBEPFSLWMLFAAGKGVGLMFGVAEPLTHY 115
QY 116 LRDVHPEPEPADRLQAILPPLGLGLVALLRMNTFAAFVYLAELMGSVLSLMFGFANEI 175
DB 116 LSDIYGAHEHQOQAL-----HTLFHWGHAMAV-----YGTIALALAYFGFRKL 163
QY 176 T-----KIHEAKRFYA-----LFGI-----GANISILASG-RATYVASKL 209
DB 164 PLALRSCTPYLLKDRINKIGDAIDVMAALLATLFGIITTLFGSSQALGAEIQIMISQN 223
QY 210 RASVSEGVDPWGISIRLLAMNTVSGVLVMSYVMINKVLTDPFRFYNPEEMQKKGAK 269
DB 224 SFALQGVYI-----VVMCLAVFSAISGV-----GKGVK 252
QY 270 --PKKMMDSF-----LYIDRSYIILLTLVLAIVG-ICINLIEVWKSQKLQYPMNNDY 322
DB 253 ILSEINILAFCLLFLVLSGFTLYLBSAFSDNIGNYFSNLVQLSFKT-----YAYEGEH 307
QY 323 SEPMGNFSF-----MTGVSVLIMLFGVGNVIRKFGMLTALVTPVWLLNGIVFA 374
DB 308 TSWISGTVLVYAMWCSNAPFVGLFIARISKRTIRF--IFGVIVP--SLRGILWF- 361
QY 375 LVIFRNOASGLVMPFGTTPML--AVVGAIONILSKSTKVALPDSTKEMAYIP----- 427
DB 362 -----TVGNTAVMLNDGIAAGLGEFIS-SPILLF--KFLNVLPLPPTTG 405

```

```

QY 428 -----DOE-----QKVKGAIDVVAARFGSGS-A 452
DB 406 FVSLVILLFFITSADSGIYVANNIARSDKSLASPMQAQIMWGLMSVALVIMQSGELA 465
QY 453 LIQOGLLVTCGSGIGAMTPYLAIVLLFIATIM-LVS-----ATKLN 491
DB 466 NLQMTILIVA-----LPFALLMLVMQFSLMKGLIAKKVFSTKVN 505

RESULT 39
GREECK
nitrite extrusion protein - Escherichia coli (strain K-12)
N/Alternate names: nitrite facilitator
C/Species: Escherichia coli
C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 01-Mar-2002
C/Accession: S05239; D64869
R/Noji, S.; Nohno, T.; Saito, T.; Taniguchi, S.
FEBS Lett. 252, 139-143, 1989
A/Title: The narX gene product participates in nitrate transport induced in Escherichia
A/Reference number: S05239; MUID:6938707; PMID:2668029
A/Accession: S05239
A/Molecule type: DNA
A/Residues: 1-463 <NCU>
A/Cross-references: EMBL:X15996; NID:942089; PIDN:CAA34126.1; PID:942091
A/Experimental source: strain K-12
R/Bibbner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.D.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: D64869
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-463 <BLAT>
A/Cross-references: GB:AB000220; GB:U00096; NID:91787467; PIDN:AAC74307.1; PID:91787475;
A/Experimental source: strain K-12, substrain MG1655
C/Genetics: narX
A/Map position: 27 min
C/Function:
A/Description: promotes a rapid rate of anaerobic nitrate reduction and the continuous e
A/Note: induction by nitrate
C/Superfamily: nitrate transport protein narX
C/Keywords: nitrate assimilation; nitrate transport; transmembrane protein
F/44-60/Domain: transmembrane #status predicted <TM1>
F/74-90/Domain: transmembrane #status predicted <TM2>
F/107-123/Domain: transmembrane #status predicted <TM3>
F/132-148/Domain: transmembrane #status predicted <TM4>
F/180-196/Domain: transmembrane #status predicted <TM5>
F/218-234/Domain: transmembrane #status predicted <TM6>
F/256-272/Domain: transmembrane #status predicted <TM7>
F/320-336/Domain: transmembrane #status predicted <TM8>
F/348-364/Domain: transmembrane #status predicted <TM9>
F/405-421/Domain: transmembrane #status predicted <TM10>
F/436-452/Domain: transmembrane #status predicted <TM11>

Query Match 5.0%; Score 132.5; DB 1; Length 463;
Best Local Similarity 22.3%; Pred. No. 0.014;
Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;

QY 69 MLVVPICAILFMILYAKLSNI-LSKQALFYAVGPPFLIFPALFPIVIVPLRDVLPHTPEAD 127
DB 39 WISVPCILLAFPCWMLFSAVAVMDPKGFNPTDQLMFLALPSVSGALRV--PYSF-- 94
QY 128 RLQAILPPGLGLVAIIL--RNWTFAPFYVLAELMGSVLSLMFGFANEITKHEAKRFY 185
DB 95 -----WPIFGGRWT--AFST-----GLIILICVWLGF-----VQDSTPY 130
QY 186 ALPEIGANISLLASGRAIVWASKLRASVS-----EGVDPGISIRLLAM 230
DB 131 SVFTI-----ISLLGFAGANAFSSM-ANISFFPKQKOGALGILNGIGVSVMLQVAP 186

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:07:42 / Search time 53 Seconds

(without alignments)
1542.343 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630
Sequence: 1 MTKTEKPEFKLRSLFMPDIH.....AQSLKEQEVAGQSDAPASS 515

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 100 summaries

A.GeneSeq.190un03:*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	2630	100.0	515	21	AAV90265
2	2617	99.5	524	20	AAV34951
3	2141	81.4	529	20	AAV37153
4	1330.5	50.6	523	21	AAAG13017
5	1330.5	50.6	618	21	AAAG13016
6	1330.5	50.6	655	21	AAAG13015
7	1220.5	46.4	587	21	AAV50990
8	1156.5	44.0	569	21	AAV50991
9	1054	40.1	498	21	AAV50992

10	1025.5	39.0	540	22	AAV74452
11	1025.5	39.0	551	20	AAV55235
12	939.5	35.7	535	20	AAV37729
13	162	6.2	538	22	AAAG89863
14	157.5	6.0	907	20	AAV34888
15	152	5.8	420	20	AAV36963
16	144	5.5	660	21	AAV74601
17	144	5.5	660	21	AAV74602
18	144	5.5	675	23	AAU91074
19	134	5.1	660	21	AAV74600
20	134	5.1	675	24	ABP78636
21	132.5	5.0	463	22	AAU34551
22	129	4.9	462	22	AAU34568
23	125.5	4.8	713	24	ABJ19074
24	122.5	4.7	713	21	AAV58582
25	120.5	4.6	395	22	AAU38299
26	120.5	4.6	425	24	AAU36578
27	119.5	4.5	472	22	ABJ19068
28	119.5	4.5	800	23	ABP40810
29	119	4.5	517	20	AAV49633
30	118	4.5	466	22	AAU38238
31	118	4.5	694	23	ABB93711
32	117.5	4.5	430	22	AAU34182
33	117.5	4.5	744	23	ABB48123
34	117	4.4	493	23	ABU8247
35	116.5	4.4	635	24	AAE30467
36	116	4.4	410	22	AAAG90107
37	115.5	4.4	410	22	AAAB78947
38	115.5	4.4	392	22	AAE05852
39	115	4.4	708	22	AAAG91270
40	115	4.4	403	23	ABB71346
41	114	4.3	478	22	ABP39792
42	113	4.3	606	20	AAV19800
43	113	4.3	633	20	AAV19799
44	113	4.3	1095	22	AAU69878
45	113	4.3	1095	22	AAU69878
46	113	4.3	1095	22	AAU69878
47	113	4.3	1095	22	AAU69878
48	113	4.3	1095	22	AAU69878
49	113	4.3	1095	22	AAU69878
50	113	4.3	1104	21	AAV95437
51	113	4.3	1104	21	ABG61818
52	113	4.3	1681	23	AAU78573
53	112.5	4.3	527	24	ABJ37097
54	112.5	4.3	535	24	ABJ37081
55	112	4.3	446	23	ABB54869
56	112	4.3	680	22	ABB65799
57	112	4.3	680	22	ABB66097
58	111.5	4.2	388	15	AAE57032
59	111.5	4.2	471	24	ABJ19193
60	111.5	4.2	1083	22	AAAG3160
61	111	4.2	360	18	AAW14079
62	111	4.2	360	18	AAW22178
63	111	4.2	741	22	AAAG90701
64	111	4.2	1095	22	AAU69878
65	111	4.2	1095	22	AAU69878
66	111	4.2	1095	23	ABB55339
67	111	4.2	1095	24	ABU71770
68	110.5	4.2	425	24	AAU4874
69	110.5	4.2	425	22	AAAG8958
70	110.5	4.2	443	22	AAAG8958
71	110.5	4.2	476	22	AAAG91987
72	110.5	4.2	498	23	ABP77193
73	110.5	4.2	691	23	ABP69062
74	110.5	4.2	697	23	AAAB78983
75	110.5	4.2	462	23	AAAB55338
76	110	4.2	927	23	ABP74086
77	109.5	4.2	665	23	AAE21150
78	109.5	4.2	691	23	AAE22915
79	109	4.1	538	22	ABB61067
80	109	4.1	551	22	ABB67783
81	109	4.1	683	22	AAU44669
82	108.5	4.1	473	20	AAV36998

Chlamydia pneumoniae
Chlamydia pneumoniae
Chlamydia trachoma
C glutamicum prote
Amino acid sequenc
Amino acid sequenc
Neisseria meningit
Neisseria meningit
Neisseria gonorrhoe
Neisseria gonorrhoe
E. coli cellular p
E. coli cellular p
Pathogen specific
Sorangium cellulos
Salmonella typhi c
Staphylococcus aur
Staphylococcus aur
Staphylococcus epi
Wheat hexose carri
Salmonella typhi c
Herbicidally activ
Staphylococcus aur
Listeria monocytog
Salmonella typhi c
Haemophilus influe
C glutamicum prote
C glutamicum SRT
Pseudomonas stutze
C glutamicum prote
Drosophila melanog
Staphylococcus epi
B. burgdorferi ant
Prostate-tumour de
Human prostate CDN
P788P amino acid s
Human P788P protei
Prostate cancer as
Human calcium chan
Prostate cancer-as
Mouse voltage gate
OspH1 rice phospho
High-affinity phos
Lactococcus lactis
Drosophila melanog
Drosophila melanog
Serotonin receptor
Pathogen specific
C glutamicum prote
S. thermophilus exo
S. thermophilus exo
C glutamicum prote
Human prostate CDN
P788P amino acid p
Human P788P polyom
Prostate cancer as
E. coli cellular p
E. coli growth and
Corynebacterium gl
C glutamicum prote
Streptococcus poly
Human polypeptide
C. glutamicum SRT
Lactococcus lactis
Candida albicans e
Human 52991 protei
Human transporter
Drosophila melanog
Drosophila melanog
Propionibacterium
Chlamydia trachoma

83 108.5 4.1 485 24 AAB76889 N. gonorrhoeae am
84 108.5 4.1 587 22 AAU34198 Staphylococcus aur
85 108.5 4.1 596 22 AAU33748 Staphylococcus aur
86 108.5 4.1 603 22 AAU37074 Staphylococcus aur
87 108.5 4.1 604 22 AAU36932 Staphylococcus aur
88 108 4.1 422 23 AAB30076 Streptococcus poly
89 108 4.1 451 22 AAU33583 Streptococcus aeri
90 108 4.1 477 22 AAB47922 Pseudomonas faeca
91 108 4.1 868 21 AAG41508 Arabidopsis thalia
92 108 4.1 879 21 AAG41507 Arabidopsis thalia
93 108 4.1 890 21 AAG41506 Arabidopsis thalia
94 108 4.1 933 22 AAU36224 Pseudomonas aerugi
95 107.5 4.1 413 23 AAB49725 Listeria monocyog
96 107.5 4.1 436 23 AAB29086 Streptococcus poly
97 107.5 4.1 507 23 AAB48420 Listeria monocyog
98 107 4.1 527 19 AAB98268 H. pylori GHP0 335
99 107 4.1 841 21 AAB11039 S. xylosum mpf pr
100 107 4.1 1043 22 AAG92893 C glutamicum prote

ALIGNMENTS

RESULT 1

AA90265 standard; Protein; 515 AA.

AA90265;

22-SEP-2000 (first entry)

C. pneumoniae ATP/ADP translocase protein sequence.

ATP/ADP translocase; Chlamydia infection; diagnosis; therapy.

Chlamydia pneumoniae.

WO200039157-A1.

06-JUL-2000.

22-DEC-1999; 99WO-CA01224.

28-DEC-1998; 98US-0114060.

12-MAR-1999; 98US-0123967.

30-JUN-1999; 99US-0141271.

(CONN-) CONNAUGHT LAB LTD.

Muridin AD, Oomen RP, Wang J, Dunn P;

WPI; 2000-452368/39.

N-PSDB; AAA30922.

Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,

prevention and treatment of Chlamydia infection in mammals -

Claim 16; Fig 1; 81pp; English.

This sequence represents the Chlamydia pneumoniae ATP/ADP translocase

of the invention. The protein, DNA encoding it, or a vaccine containing

the DNA or protein, are useful for diagnosing, preventing or treating

Chlamydia infection. The sequences can also be used in a method for

the detection of Chlamydia infection. Primers or probes derived from the

DNA sequence are useful in diagnostic tests for detecting Chlamydia

infection.

Sequence 515 AA;

Query Match 100.0%; Score 2630; DB 21; Length 515;

Best Local Similarity 100.0%; Pred. No. 2.2e-282;

Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTEKPFKLRSLPMLPHTHELKVLPMELMPCITFNTYVLTDTKDTLIVAGPSGA 60
Db 1 MTKTEKPFKLRSLPMLPHTHELKVLPMELMPCITFNTYVLTDTKDTLIVAGPSGA 60
QY 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPLIFPALFPIYIPLRVL 120
Db 61 EAIPIKFMVLVPCAIIFMLIVAKLSNLSKQALFYAVGTPLIFPALFPIYIPLRVL 120
QY 121 HPTFADRLQAILPPGLIGVAIILNMTFAFVYLAELWGSVMSLMFMGFANETIKHE 180
Db 121 HPTFADRLQAILPPGLIGVAIILNMTFAFVYLAELWGSVMSLMFMGFANETIKHE 180
QY 181 AKRFYALFGIGANISILASGRALVWASVSEGVDPWGISLRLMAMTIVSGLVMA 240
Db 181 AKRFYALFGIGANISILASGRALVWASVSEGVDPWGISLRLMAMTIVSGLVMA 240
QY 241 SYWMINKVLTDPREFYNEEMQKGGAKPPONMKDSFLYIDRSPIYILLTLVYAGIC 300
Db 241 SYWMINKVLTDPREFYNEEMQKGGAKPPONMKDSFLYIDRSPIYILLTLVYAGIC 300
QY 301 INLEIVTWKSQKIQYPMNDYSEFMGNFSGVTVSLIMLVGNAVIRKFGMTGALV 360
Db 301 INLEIVTWKSQKIQYPMNDYSEFMGNFSGVTVSLIMLVGNAVIRKFGMTGALV 360
QY 361 TPVWVLITGIVFPAIVIRNOASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTX 420
Db 361 TPVWVLITGIVFPAIVIRNOASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTX 420
QY 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVITGSGTAMTPYLAIVLFI 480
Db 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVITGSGTAMTPYLAIVLFI 480
QY 481 AIVWSATKTKNLFLAQSALKEQEVADSDAPASS 515
Db 481 AIVWSATKTKNLFLAQSALKEQEVADSDAPASS 515

RESULT 2

AA934951 standard; Protein; 524 AA.

AA934951;

13-SEP-1999 (first entry)

Chlamydia pneumoniae transport polypeptide.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

vaccine; neutralising epitope.

Chlamydia pneumoniae.

WO9927105-A2.

03-JUN-1999.

20-NOV-1998; 98WO-IB01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(BEST) GENSET.

Griffais R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Page 871-872; Disclosure; 1912pp; English.

AA934584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 524 AA;

Query Match 99.5%; Score 2617; DB 20; Length 524;

Best Local Similarity 99.6%; Pred. No. 6, 2e-281;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPPGKRSFLMPTHTLKKVLPMLMFCITPNYTVLRDTKTLIVGAGSGA 60
 DB 10 MTKTEKPPGKRSFLMPTHTLKKVLPMLMFCITPNYTVLRDTKTLIVGAGSGA 69
 QY 61 EAIPIKFMVAVPCAIIFMLIYAKSNILSKOALFAVGTPLPFIPTIYPLRDVL 120
 DB 70 EAIPIKFMVAVPCAIIFMLIYAKSNILSKOALFAVGTPLPFIPTIYPLRDVL 129
 QY 121 HPTFPADRLQALIPGLGLVALILRNWTPAFAVLAELMGVWLSLMFGFANEITKIH 180
 DB 130 HPTFPADRLQALIPGLGLVALILRNWTPAFAVLAELMGVWLSLMFGFANEITKIH 189
 QY 181 AKRFVALFGIGANISILASGRAIWAASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 190 AKRFVALFGIGANISILASGRAIWAASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 249
 QY 241 SYWMINKNVLTDPRFNPBEMOKGKGAAPKKNMDSFLYLDSPYITLLTLVAVGIC 300
 DB 250 SYWMINKNVLTDPRFNPBEMOKGKGAAPKKNMDSFLYLDSPYITLLTLVAVGIC 309
 QY 301 INLIEVTWKSQKLOQYPMNDYSEPMGNFSFWTGVSVLIMLVGNAVIRKFGMLTGALV 360
 DB 310 INLIEVTWKSQKLOQYPMNDYSEPMGNFSFWTGVSVLIMLVGNAVIRKFGMLTGALV 369
 QY 361 TPVWVLGTGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTYALPDSK 420
 DB 370 TPVWVLGTGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTYALPDSK 429
 QY 421 EMAYIPLDQOKVKGKAIDVVAARFGSGGALIOQGLVTCGSI GANTPYLAVALPFI 480
 DB 430 EMAYIPLDQOKVKGKAIDVVAARFGSGGALIOQGLVTCGSI GANTPYLAVALPFI 489
 QY 481 AIMLVSATKLNKFLAQSALKEQVADSDAPASS 515
 DB 490 AIMLVSATKLNKFLAQSALKEQVADSDAPASS 524

RESULT 3

AAY37153 standard; Protein; 529 AA.

AAY37153;

07-OCT-1999 (first entry)

Chlamydia trachomatis transport protein.

Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;
 nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

bartholinitis; pneumopathy; venereal lymphogranulomatosis.

Chlamydia trachomatis.

OS Chlamydia trachomatis.
 XX MO9928475-A2.
 XX PN

PD 10-UTN-1999.
 XX 27-NOV-1998; 98WO-IB01939.
 PF 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015044.
 PR 17-DEC-1997; 97FR-0016034.
 XX (GIST) GENSET.
 PA Griffiths R;
 XX WPI, 1999-371125/31.
 DR Genome sequence of Chlamydia trachomatis

PT Disclosure; Page 935-936; 1755pp; English.

PS AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perinephritis, bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX Sequence 529 AA;

Query Match 81.4%; Score 2141; DB 20; Length 529;

Best Local Similarity 77.9%; Pred. No. 4e-228;
 Matches 409; Conservative 52; Mismatches 50; Indels 14; Gaps 3;

QY 1 MTKTEKPPGKRSFLMPTHTLKKVLPMLMFCITPNYTVLRDTKTLIVGAGSGA 60
 DB 2 MTKTEKPPGKRSFLMPTHTLKKVLPMLMFCITPNYTVLRDTKTLIVGAGSGA 61
 QY 61 EAIPIKFMVAVPCAIIFMLIYAKSNILSKOALFAVGTPLPFIPTIYPLRDVL 120
 DB 62 EAIPIKFMVAVPCAIIFMLIYAKSNILSKOALFAVGTPLPFIPTIYPLRDVL 121
 QY 121 HPTFPADRLQALIPGLGLVALILRNWTPAFAVLAELMGVWLSLMFGFANEITKIH 180
 DB 122 HPTFPADRLQALIPGLGLVALILRNWTPAFAVLAELMGVWLSLMFGFANEITKIH 181
 QY 181 AKRFVALFGIGANISILASGRAIWAASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 240
 DB 182 AKRFVALFGIGANISILASGRAIWAASKIRASVSEGVDPWGISLRLLMAMTIVSGLVMA 241
 QY 241 SYWMINKNVLTDPRFNPBEMOKGKGAAPKKNMDSFLYLDSPYITLLTLVAVGIC 300
 DB 242 SYWMINKNVLTDPRFNPBEMOKGKGAAPKKNMDSFLYLDSPYITLLTLVAVGIC 299
 QY 301 INLIEVTWKSQKLOQYPMNDYSEPMGNFSFWTGVSVLIMLVGNAVIRKFGMLTGALV 360
 DB 300 INLIEVTWKSQKLOQYPMNDYSEPMGNFSFWTGVSVLIMLVGNAVIRKFGMLTGALV 359
 QY 361 TPVWVLGTGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTYALPDSK 420
 DB 360 TPVWVLGTGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTYALPDSK 419
 QY 421 EMAYIPLDQOKVKGKAIDVVAARFGSGGALIOQGLVTCGSI GANTPYLAVALPFI 480
 DB 420 EMAYIPLDQOKVKGKAIDVVAARFGSGGALIOQGLVTCGSI GANTPYLAVALPFI 479
 QY 481 AIMLVSATKLNKFLAQSALKEQVADSDAPASS 513
 DB 480 AIMLVSATKLNKFLAQSALKEQVADSDAPASS 524


```
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154033.
PR 20-SEP-1999; 99US-0154179.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155658.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158023.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161355.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 50.6%; Score 1330.5; DB 21; Length 523;

Best Local Similarity 50.9%; Pred. No. 3.2e-138;

Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

```
QY 16 LMPHTHEIKKLVLPNFMFPCTTENTYVLRDPTKDTLIIYGAPSSGAEALPFIKFWLVVCA 75
DB 7 IFGEVTLTKKIVPGLMEFCILFNYTIRDKDVLVYAKGSAEIIPELKTWNLPMA 66
QY 76 IIFMFIYAKSNILSKOLFVAVGTPELIFPALFPTVIVPLRDLVAPTEPFAADRLQAIIP 135
DB 67 IGFMLTYTKLSVNLSSKALFVIVPFIYVFCAGFGWVPLSNLHPEALADKLATIGP 126
QY 136 GLLGLVAILRNNTFAFYVLAELWGSVMLSMFMGPFANEITKIHEAKRPYALFGIGANIS 195
DB 127 REMGPLAIRIWSFCIFYVMALMGSSVVSVLFWMGFANOITTVDBAKKFFYPLFGIGANVA 186
```

```
QY 196 LLASGRAIVWASKLKASVSEGVDPMGISRLILMAMTIVSGVLVMASSWINKVLTDPRE 255
DB 187 LIFSGRTVKYFSSMKRKLGGVDGMASVLSKAMSIYVGGLACFLYWNVR----- 238
QY 256 YNPEEMQKGGKAKPKKMMKDSFLYDRSPYILLTLVIAVGCINLIEVTKSQKIQ 315
DB 239 YVPLPTRSKKKKVPQGTWESLKFVSSPYIRDLATLVAYGISINLVEVTWKSXLQAQ 298
QY 316 YPMNNDYSEPMGNPSFTGTAVSLIMLFVGANVIRKRGWLTGALVPMVWLGIYFEAL 375
DB 299 FSPSENEYSAPMGDSFTGTATFTMM-LISQYVFKKYGWVAAKITPTVILLTGVAFFSL 357
QY 376 VIFPNQASGVAMFGTPTMLAVVGAIQNILSKSTYKALFDS TKEMAVYPLDQEQVKYG 435
DB 358 ILFGPPAPVAVKLGMTPLAAVYVGAIQNIFSKSAKSLFDPCKEMAYITPDEDTVKG 417
QY 436 KAAIDVVAARFGKSGGALLIOGGLVIGSIGAMTPYIAVILLPTIAIWTVSATKL---N 491
DB 418 KAAIDVVCNPLGKSGGALLIOGFMLTFGSLANSTPYIGVILLGIVTAMLAASKLLEGQFN 477
QY 492 KLFLAQGALKQE-----VAQEDSAPA 513
DB 478 TLMSEELEREMERASSVKIPVVSQED-APS 507

RESULT 5
AAG13016
ID AAG13016 standard; Protein; 618 AA.
XX
XX AAG13016;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 12352.
DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 12352.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0144941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136787.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143524.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144362.
PR 20-JUL-1999; 99US-0144844.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149375.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154019.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158359.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

[illegible]

PR	05-MAR-1999	99US-0123180
PR	09-MAR-1999	99US-0123148
PR	23-MAR-1999	99US-0125788
PR	23-MAR-1999	99US-0126764
PR	29-MAR-1999	99US-0126762
PR	01-APR-1999	99US-0127362
PR	06-APR-1999	99US-0128324
PR	08-APR-1999	99US-0128714
PR	16-APR-1999	99US-0129845
PR	19-APR-1999	99US-0130077
PR	21-APR-1999	99US-0130449
PR	23-APR-1999	99US-0130510
PR	23-APR-1999	99US-0130891
PR	30-APR-1999	99US-0131449
PR	30-APR-1999	99US-0132048
PR	04-MAY-1999	99US-0132407
PR	04-MAY-1999	99US-0132484
PR	05-MAY-1999	99US-0132485
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0132463
PR	11-MAY-1999	99US-0134456
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	14-MAY-1999	99US-0134270
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134768
PR	18-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135523
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136792
PR	28-MAY-1999	99US-0136982
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139452
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139458
PR	18-JUN-1999	99US-0139459
PR	18-JUN-1999	99US-0139460
PR	18-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139750
PR	21-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139869
PR	23-JUN-1999	99US-0140053
PR	23-JUN-1999	99US-0140054
PR	24-JUN-1999	99US-0140055
PR	28-JUN-1999	99US-0140823
PR	29-JUN-1999	99US-0140591
PR	30-JUN-1999	99US-0141867
PR	01-JUL-1999	99US-0141842
PR	02-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142820
PR	12-JUL-1999	99US-0142877
PR	13-JUL-1999	99US-0143542

```
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.6%; Score 1330.5; DB 21; Length 655;
Best Local Similarity 50.9%; Pred. No. 4,4e-138;
Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;

QY 16 LMPHTHEKLVVPMFLMFCITRTNYVLRPTKOTLVGABGSAEMIPFKFLVVPKA 75
DB 139 IFGEVVTLLKKIVPLGLMFCILFNNTILRDTKVLVVTAKGSAEIIPLKTVNLPM 198
QY 76 IIFMLIYAKLSNISKOALFYAVGTPFLIPALFPVIVPRLDVLHPTBFDRLQALPP 135
DB 199 IGFMLIYAKLSNISKOALFYAVGTPFLIPALFPVIVPRLDVLHPTBFDRLQALPP 258
QY 136 GLIGLIVAILRNWTPAAYVLAELMGVYMLSMFMFANEITKIHAKRFYALFGIGANIS 195
DB 259 RFMGPLAIMRIMWSPCLFYVAABELMGVYVSLFMFGFANQITTVDEAKKFFPLFOLGANVA 318
QY 196 LLASGRAIWAASKIRASVSEGVDPWGISLKLMAAMTIVSGIULMASIYMWINKVLTDPFR 255
DB 319 LIFSGRTVYKFFSNRNKVLGPGVDGMAVSLKAMMSIVGMGLAICFLVMWNR----- 370
QY 256 YNPEMOKKKGAKPKKNMKSFLYLDSPYILLTLVAVGICVILEVTWMSOKLQ 315
DB 371 YVLPPTSKKKKKVPPQGMTHESLKFVSSPIRDLATLVAVGISINLVEYTKSKLKAQ 430
QY 316 YPNNDYSEFMGNFSFTGVSVLIMLFVGANVIRKFGMLTGAIVTPVWVLTLGIVEPAL 375
DB 431 FSPBNYSAPMGDRSTGTGATFTMML-LSQYVFKKYGWGAAKITPTVLLTGVAFFSL 489
QY 376 VIFPNQASGLVAFGTTPPLMLAVVGAIONILSKSTYALFDSITKEAAYIPLDOEQVKG 435
DB 490 ILFGPPAPLVAKLGMPPLAAVYVGAIONIFSKSAKYSFLDCKEKAAVYIPLDEDTVKKG 549
QY 436 KKAIDVVAARFGSGGALLIOGGLLVICSGANTPYLAVLTLFIATMLVSATKL-----N 491
DB 550 KKAIDVVCNPLGSGGALLIOGGLLVICSGANTPYLAVLTLFIATMLVSATKL-----N 609
QY 492 KLFLAOSALKOE-----VAODESAPA 513
DB 610 TLMSEELBREMERASVKKIPVVSQED-APS 639
```


KM bronchitis; sinusitis; atherosclerosis; asthma; vaccine; antimicrobial;
 KM antibacterial; diagnosis.
 XX Chlamydia pneumoniae.
 OS
 XX WO200121803-A1.
 PN
 XX 29-MAR-2001.
 PD
 XX 15-SEP-2000; 2000WO-CA01087.
 PF
 XX 17-SEP-1999; 99US-0154326.
 PR
 XX (AVET) AVENTIS PASTEUR LTD.
 PA
 XX Murdin AD, Oomen RP, Wang J, Dunn P;
 PI
 XX WPI; 2001-316102/33.
 DR
 XX N-PSDB; AAF84486.
 XX
 PT New Npt2cp (ADP/ATP translocase) polypeptides and nucleic acids
 PT encoding the polypeptides useful for treating, preventing or diagnosing
 PT Chlamydia infections, particularly infections caused by Chlamydia
 PT pneumoniae
 PS
 XX Claim 16; Fig 1; 79pp; English.
 XX
 CC This sequence represents the Chlamydia pneumoniae Npt2cp (ADP/ATP
 CC translocase) protein. Chlamydia pneumoniae Npt2cp nucleotides and
 CC proteins may be used for preventing, treating or diagnosing
 CC chlamydial infections, particularly those caused by Chlamydia
 CC pneumoniae. Chlamydia pneumoniae can be responsible for both upper and
 CC lower respiratory tract infections. It is the third most common cause of
 CC community acquired pneumonia, and can also cause bronchitis and
 CC sinusitis. It is also linked to diseases other than respiratory tract
 CC infections, being associated with atherosclerosis and asthma.
 CC Npt2cp nucleotides may be used for the recombinant production of the
 CC protein, as genetic vaccines, in the construction of vaccine vectors
 CC such as poxviruses, and in the construction of attenuated Chlamydia
 CC strains which can over-express an Npt2cp nucleotide, or which can
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used
 CC in vaccine compositions, and may also be used as diagnostic reagents for
 CC detecting the presence of anti-Chlamydia antibodies in a sample.
 CC
 XX
 SQ Sequence 540 AA;
 Query Match 39.0%; Score 1025.5; DB 22; Length 540;
 Best Local Similarity 39.4%; Pred. No. 2.3e-104;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;
 QY 1 MTKTEKPEFGKLRSLFPIHTHELKVVLPMLFECITFNVTYVLRDTKDTLIVGAPGSGA 60
 DB 1 MOSSEKPEPSRLRAYLCPIYKSEBSKFPVPLFLAFVFGNYCLKMKMDTLIVGSDGA 60
 QY 61 EAIPIFKFPLVVPACAIIFMLIYAKLSNLSKQALFAVAGTPEPLPALEPTIYPLRDVL 120
 DB 61 EYIPFLKVGIVGAVITVWVGWGLGSRPRDTVFCFMAALGFFLEFAVITTYGDSL 120
 QY 121 HPTFPAQLQALIPGLGLVAILRNWTPAFAVLAELWGSYMLSMFWGANETIKIHE 180
 DB 121 HNSLADKLQELPQGLRFGFIWVRYSYIVVMSLSVVLGMLFGLANQTTTITE 180
 QY 181 AGRFVALFSGIGANISILASGRATWASKR-ASVSGVDPW-GISRLIMANTIVSGVL 238
 DB 181 AGRFVALFSGIGANISILASGRATWASKR-ASVSGVDPW-GISRLIMANTIVSGVL 238
 QY 239 MASVWINKV--LTPRFYNEEMOKGKGA-----PKPMKMSFYLDRSPYI 287
 DB 240 I-----WLYRIRHITLDTISIPSRRLAEGAAITANLKKKKPKAKARLFLHLQSKRL 295
 QY 288 LLLTLIVAYGICINILEVTKSOLKQYPNNDVSEFMGNFSFWTGVSVLIMLFGVGN 347
 DB 296 LGLAITIVLSYNVILHFEVVMKQVQVQISVSHVEFNGYSRITTLIGVSVLAVALTNGO 355

QY 348 VIRKGMULTGALVPMVILNGIVFPALVI-FRNOASGLVMPGTTPLMLAVVGAIONI 406
 DB 356 CIRKGMVTVGALVPLVWLVSGLLFPGTIFPAKADISIFGVLGMPALAAAMTGMQNV 415
 QY 407 LSKSTKVALPDSTEMAYIPLDOEKVKGAIDVVARFPGSGALIOGLLVICSGIG 466
 DB 416 LSRGKTFPFQTKEMAFIPLSPEDKHKGAIDGVSRIGKSGSLIYQGLLVIFSSVA 475
 QY 467 AMTPYLAVILFLTAIWLVSATKANKLFLAQA-----LK-----RQVNAQE 508
 DB 476 ASLNVIALVLLIIMVWIAVAVYIGKEYSRADAVALTKPKPESSSIYREAGSEVQE 535
 QY 509 DSA 511
 DB 536 ENA 538
 RESULT 11
 ID AAY35235
 AAY35235 standard; Protein: 551 AA.
 AC AAY35235;
 XX
 DT 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae transport polypeptide.
 XX
 DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.
 OS Chlamydia pneumoniae.
 XX
 EN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-1B01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GBST) GENSET.
 XX
 PI Griffiths R;
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Page 1080-1081; Disclosure; 1912pp; English.
 XX
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory diseases such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX
 SQ Sequence 551 AA;
 Query Match 39.0%; Score 1025.5; DB 20; Length 551;
 Best Local Similarity 39.4%; Pred. No. 2.3e-104;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;
 QY 1 MTKTEKPEFGKLRSLFPIHTHELKVVLPMLFECITFNVTYVLRDTKDTLIVGAPGSGA 60
 DB 1 MOSSEKPEPSRLRAYLCPIYKSEBSKFPVPLFLAFVFGNYCLKMKMDTLIVGSDGA 60

```

Db      12  MGSSEVFPGRRLRAVLCPIYKSEFSKVPFLFLAFVGVNCLLKKMKDTLVI VSGDAG 71
QY      61  EALPFLKFLVWCALITFMILYAKLSNLSKQALFVAVGPFLIPALPFTVYIPRLDV 120
Db      72  EVLPFLKFLVWCALITFMILYAKLSNLSKQALFVAVGPFLIPALPFTVYIPRLDV 131
QY      121  HPTFADRLQALIPGGLGLVALILRWTFAPFVYLAELMGSVMSLMFQFANEITKHE 180
Db      132  HMLSLADKQGLPGLPGRGFIWVRVWSYIYVMSLMSVLSLMLFQNLQITTFE 191
QY      181  AKRFYALFGIGANISLIASGRAIYVASKLR-ASVSEGVDPW-GISRLIMATVYSGVL 238
Db      192  AGRFYALINTGLMSSICAGEISYMMGKOTFVAVSPACDSMHSVMNLITLIT-CSGLIM 250
QY      239  MASVYMINKNV--LTDRFVNPBEOMKGGKA-----KPKMMKOSFLYLDSPYI 287
Db      251  I---WYRRIHHLITDTSTIPFSRRVLAEGAAFTANLKEKKRKAKARMLFHLIQSRYL 306
QY      288  LLTLTLVAVGICINIEVTWKSQKLQYENMNDYSEFMGNFSPMTGVSVLIMLPVGN 347
Db      307  LGLAIVLSNLYIHLEFVVMKQVSOIYSSHYEFNGYSRITTLIGVSVLAVALITQ 366
QY      348  VIRKFMVLGALVTPVAVLTGIVFPALVI-FRNQASGLVAMEGTTPLMLAVVGAIQNI 406
Db      367  CIRKMGMTVGAIVTPVAVLTGIVFPALVI-FRNQASGLVAMEGTTPLMLAVVGAIQNI 426
QY      407  LSKSTKVALPDSRKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIG 466
Db      427  LSKSTKVALPDSRKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIG 486
QY      467  AMTPYLAIVLFLITAIWVSATKLNKFLAQA-----LK-----EOEVAQE 508
Db      487  ASLNVALVLLVIMVAVVAVIGKEYYSRAADAVALIKQPREPSSIVREAGSVBE 546
QY      509  DSA 511
Db      547  EMA 549

RESULT 12
AA37729
ID  AA37729 standard; Protein; 535 AA.
AC  AA37729;
XX
XX  07-OCT-1999 (first entry)
DE  Chlamydia trachomatis transport protein.
XX
XX  Vaccine, eye disease; conventional trachoma; nonendemic trachoma;
KW  paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;
KW  nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW  bartolinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS  Chlamydia trachomatis.
XX
XX  WO928475-A2.
XX
XX  10-JUN-1999.
XX
XX  27-NOV-1998; 98WO-1B01939.
XX
XX  04-NOV-1998; 98US-0107077.
XX  28-NOV-1997; 97FR-0015041.
XX  17-DEC-1997; 97FR-0016034.
XX
XX  (GSEST ) GENSEST.
XX  PA
XX  Grifffals R.
XX  PI
XX  MPI, 1999-371125/31.
XX
XX  Genome sequence of Chlamydia trachomatis
PT

```

```

XX      XX  Disclosure; Page 1329-1330; 1755bp; English.
PS      CC  AA37729-1330 are encoded by open reading frames (ORFs) of the genome
XX      CC  of Chlamydia trachomatis (see A201425). The polypeptides can be used as
XX      CC  vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX      CC  can also be used to control growth of the microorganism. Chlamydia
XX      CC  trachomatis is responsible for a large number of diseases, e.g. eye
XX      CC  diseases such as conventional trachoma, nonendemic trachoma,
XX      CC  paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX      CC  nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX      CC  perhepatitis, bartolinitis; pneumopathy in breast feeding infants;
XX      CC  and venereal lymphogranulomatosis. The polypeptides of the invention
XX      CC  may be of use in treating these diseases.
SQ      XX  Sequence 535 AA;
        XX  Query Match 35.7%; Score 939.5; DB 20; Length 535;
        XX  Best Local Similarity 37.2%; Pred. No. 7.8e-95;
        XX  Matches 195; Conservative 113; Mismatches 195; Indels 21; Gaps 6;

QY      8  PPKIRSFMLPIHHEKLVLPFLMFCITENTYVLRDTKOTLIVGAPSGAALIPK 67
Db      2  PFRSGDXFFPIYSEFSKFIPLFLAFVGVNVALTKTQDSIVLVGSRAGAIVIPFK 61
QY      68  FVLVPCALIFMLYAKLSNLSKQALFYANGTFLIPALPFTVYIPRLDVLPTEPAD 127
Db      62  VMLVPCALIVTMVYAKLSNLSKQALFYANGTFLIPALPFTVYIPRLDVLPTEPAD 121
QY      128  RLQALPPLGLGLVALILRWTFAPFVYLAELMGSVMSLMFQFANEITKHEAKRFYAL 187
Db      122  KLQSLPPLGGRGFVVMQVYSLSLYVMSLMSVLSLTFMGVANHITSVREAGRYAL 181
QY      188  FGIGANISLIASGRAIYVASKLR-ASVSEGVDPGIGISRLIMATVYSGVLVASYWIN 246
Db      182  INTGLMSSVFAGEVSLMGRSPVIAFPMAVDPMHEMLNITLITVLAGVITLYOKLD 241
QY      247  KNLVLPDRFVNP-----EMQKGGKGAKKPKMMKDSFLYLDSPYIILLTLVIAV 297
Db      242  R-LMDETSMLEELAAEMSVQQLKKEKSKAK--AKSLPALLRSRYLGLIADVLSY 297
QY      298  GICINLIEVTWKSQKLQYENMNDYSEFMGNFSPMTGVSVLIMLPVGNVIRKFWL 357
Db      298  NLVILHLEFVVMKQVSOIYSSHYEFNGYSRITTLIGVSVLAVALITQIRMGWTVG 357
QY      358  ALVTPVAVLTGIVFPALVI-FRNQASGLVAMEGTTPLMLAVVGAIQNLSTKVALP 416
Db      358  ALVTPVAVLTGIVFPALVI-FRNQASGLVAMEGTTPLMLAVVGAIQNLSTKVALP 417
QY      417  DSTKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIGAMPYLAIV 476
Db      418  DSTKEMAYIPLDOEKVKGAIDVVAARFGKSGGLIOGGLVITGSGIGAMPYLAIV 477
QY      477  LFIITAIWVSATKLNKFLAQA-----LKOEVAQEDSAPAS 514
Db      478  LFIITAIWVSATKLNKFLAQA-----LKOEVAQEDSAPAS 521

RESULT 13
AA37729
ID  AA37729 standard; Protein; 538 AA.
AC  AA37729;
XX
XX  26-SEP-2001 (first entry)
DE  C glutamicum protein fragment SEQ ID NO: 3617.
XX
XX  Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW  organic acid synthesis.
XX
XX  Corynebacterium glutamicum.
XX
XX

```


QY 115 PLRDLVHTEPADRLQALIPGLGLVAILRWTFPAFYVLAELMGVMS-LMPWGAN 173
 Db 110 -----AIDPP-----RSLPLF--FYRIY-IMSLTILSYTSFWGFD 141
 QY 174 EITKHEAKRFYALPGIGANISLASGRAIVASFLRASVSGVDPMGISRLMAMTIV 233
 Db 142 QPFLIDQGRHFCIN-----AIFLGDAI--GSGIISLVHTTIGIQLILFTRALVLT 194
 QY 234 SGLVIMASYWYMNKV--LTPDR--FYNPEMOKKKAKPKMNKDSFLYIDRSFYIL- 288
 Db 195 FPIV-----FYVSKSLKSLSDHDLFID-----TGHPPSLKALKLCFYDKTYFYLL 241
 QY 289 ----LITLLIAYAGCINLIEVTKWSQKLOYPNNNDV--EFMGNFSPFWGVSVLML 342
 Db 242 CFYFMQLALAI-----TEFNYSKIFEPISKEEFELVAHIGKSLMISLGMICAL 294
 QY 343 FVGGNVIRKFGMLTGALVTPVWVLLTGIVFFALVIFRNQASGLVAMFGTTPMLAVVGA 402
 Db 295 FAYSRIVKRLG-----VNNIILRAPLCFLS-----LFLWTFKTTLSIAVLA 336
 QY 403 IONILSKSTKXALPDSKEMAY--IPDQOKVKGAIDVVAARFGKSGALLIQGGLV 460
 Db 337 M--VVRGVTYALDDNNLQLIYGV-----NKIRNQ--IRIVSEFISPIGMLVMS--L 385
 QY 461 ICGSIGAMTPYLAIVLLFIILMIVSATKMLFLAQSAK 501
 Db 386 VC--FLSSQOYVFCILISLIATILVC--LVRSYAKALK 421

RESULT 15

AA36963
 ID AAY36963 standard; Protein; 420 AA.

XX AAY36963;

DT 07-OCT-1999 (first entry)

XX Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perithenarthritis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX MO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97ER-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GENSET) GENSET.

XX Griffais R;

XX WPI, 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 807-809; 1755pp; English.

XX AAY36963-437949 are encoded by open reading frames (ORFs) of the genome
 of Chlamydia trachomatis (see AAY01425). The polypeptides can be used as
 vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 can also be used to control growth of the microorganism. Chlamydia
 trachomatis is responsible for a large number of diseases, e.g. eye
 diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC Bartholinitis, pneumopathy, lymphogranulomatosis. The polypeptides of the invention
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX Sequence 420 AA;

Query Match 5.8%; Score 152; DB 20; Length 420;

Best Local Similarity 20.3%; Pred. No. 1.2e-07;

Matches 95; Conservative 74; Mismatches 170; Indels 128; Gaps 20;

QY 7 KPFGLRS--FLMFIHTEKLVLPFMFECIFENYVLRDTQTLIVAPGS----- 58

Db 15 KGSGLRALFLSLRGEKRAL-LELLGLI-----MSVACYGLAIGES 60

QY 59 -----GAELPF-----IKFVLVPCAIIMELIYAKLSNLKQALFYAVGTPELI--- 104

Db 61 XFLBEIGAKLPFAYLGSFPL---CFISCLILYNLSRKKEVSPRALFLSFICVLICNLY 117

QY 105 -FVALPFTVYPLRDVHPTPEFADRLQALIPGLGLVAILRWTFPAFYVLAELMG-SV 162

Db 118 LFWL-----AIH-----KVSQTPPL-----YRL--IWGLTI 145

QY 163 MLSLMPWGANETIKHEAKRFYALPGIGANISLASGRAIVAMSK-LRASVSGVDPWG 221

Db 146 LCVANFMGFIIDQFNIDAKRHFCIF-----NAITFGDPLGARIVNQIYLG 193

QY 222 ISRLMAMTIVSGLVIMASYWYMNKVLTDP-----FYNPEMOKKKAKPKMNK 275

Db 194 AELLALFIVITFPLVHYISSLSKESEDHDLFDYGYPSSTQ-----TLKCLK 247

QY 276 DSFLYIDRSFYILLTLIYAGCINLIEVTKWSQKLOYPNN--DYSEFMNFSFWT 333

Db 248 DKYTFYLVSEFYLMQLLVFT-----EFNYLKIIPDAFGAETCELTEFTKYSWI 299

QY 334 GVVSVLMLFVGGNVIRKFGMLTGALVTPVWVLLTGIVFFALVIFRNQASGLVAMFGTTP 393

Db 300 SLGNMFPALPAYSRYVTRFG-----INNIIILAPLCF-----PSLFCWS 339

QY 394 LMLAVGAIQNLISKSTKXALPDSKEMAYIPLDQOKVKGKAID 440

Db 340 IKTSVFAFMGTARBEGLAVALDDNNLQLIYGPINKIRNOVMAIE 386

RESULT 16
 AAY74601
 ID AAY74601 standard; Protein; 660 AA.

XX AAY74601;

DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 163 protein sequence SEQ ID NO:676.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX MO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUN-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,
 PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AA253363.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 461; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA257941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX
 SQ Sequence 660 AA;
 Query Match 5.5%; Score 144; DB 21; Length 660;
 Best Local Similarity 20.7%; Pred. NO. 1.9e-06;
 Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;
 QY 32 LMFCTITNTYVLTAKDTLVGAPG-----SGEAIPIKE-MLVVPALIFM----- 79
 DB 7 LFFVCLV-----VLVLTVPDOVMLEDAKEVIFTEFSWPFYVLTISIFLPLLI 56
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFTVYPLRDVHPTEF 125
 DB 57 LSVSLGNIRIGRDDDEPFGFLSLAMFPAAGMGVGLMFQVAPLHMYSDIAGTPE 116
 QY 126 ADRLQALIPGILGLVAILRMTFAFYVLAELMGVSLMFWGFANEITKIHAKR-- 183
 DB 117 HRQOQALL-----HTVFHMGVAMSV-----YGTIALALAVPGR---YKLPLALNSC 161
 QY 184 FYAL-----FGIGANI-SILASGRAIV-----WASKLRASVE--GVDPWGISRL 227
 DB 162 FYPLIKERKISGFGDAIDIMALLATFFGIITLFGASQGLGAGDEMGIKENSFSVQL 221
 QY 228 MAMTIVSGVLMASVWINKVLTDPFRYNEPMOKKKGAK--PKNMKCSFL-----YL 281
 DB 222 IIAAWMSLAIVSA-----ISGVGGVYVSELNLGLAFLFFVL 261
 QY 282 DRSPYILLTLTLVAVGICI-----NLIEVTWKSQKLQYPMNDYSEFMGNFSF----- 331
 DB 262 AAGPFVYLLS-----AFGDNIGNYLGNLVRLSKT-----YAKERHKWFESWTLYWAM 312
 QY 332 ---WTGVSVLIMLFVGVGNVIRKFGMLTGALVTPVAVLLTGIVFALVI FNNQASGLVAM 388
 DB 313 WCSNAPFVGLFIARISKRTIRF--VFGLIIPG-----LFGVILMP-----TV 354
 QY 369 FGTTFMLL-AVYVGAIONILSKSTKVALFEDSTKEMAYIPLDQEQV----- 433
 DB 365 FGNIAIMNDGVAGNLEKRTSSPTLLF--KFNINYPPLPLISIVSLVLSLFFVTS 411
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIOGGLVTCSSIGAM 468
 DB 412 DSGIVLVNNTSRDGLAPRWQAVMWGVLMGSAVAVLMSRSG-----LGNLSM 461

QY 469 T-----PYAVILLFIATIML-VSA-----TKLN-----KLFASQALKEQ 503
 DB 462 TLIVSLPPLMLTMCSTLMKGLSADKKYFETRVNPTSVFWTGSKMKERLVQIMSQTOEQ 521
 QY 504 EV---AQEDSAPA 513
 DB 522 DILKFLKQTASPA 534
 RESULT 17
 AA274602
 ID AA274602 standard; Protein; 660 AA.
 XX
 AC AA274602;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 163 protein sequence SEQ ID NO:678.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 OS Neisseria meningitidis.
 XX
 FM MO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PE 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,
 PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
 PI Tettelin H, Venter JC;
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253364.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS Claim 2; Page 461; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA257941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX
 SQ Sequence 660 AA;
 Query Match 5.5%; Score 144; DB 21; Length 660;

Best Local Similarity 20.7%; Pred. No. 1.9e-06;
Matches 127; Conservative 99; Mismatches 171; Indels 216; Gaps 34;

```

OY 32 LMFECITNTYVLRDTKDTLVAGP-----SGAEAIPIKPF-WLVVPCAIITM----- 79
DB 7 LFFVVCVAV-----VLVTVPDQVMWMDRAKEVITFEFSWFVLTFSIFGLFLLI 56
OY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFTVYIPLRDVLPHTPEF 125
DB 57 LSVSSLGNIRLGRDVEDPPEFGFLSWLMLFAAGVGVLMEFGVAEPLMHVPSDITATPE 116
OY 126 ADRLQALIPGGLGLVALIRNMTFAFYVLAELMGVSLMFMGFANEITKIHAKR-- 183
DB 117 HROQOALL-----HTVHHGVHAMS-----YGTIALALAYRFR--YKLPLALRSC 161
OY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVE--GVDPWGISLRLL 227
DB 162 FYPLKEKISGRFGDAIDIMALLATFFGIITTLGFGASQLGAGLQEIQWIAENSFSVQVL 221
OY 228 MAMTIVSGVLVMASTWYINKVLTDPFRYPNEBQKKGAK--PKNMKDSFL-----YL 281
DB 222 IIAVMSLAVVSA-----ISGVGKGVVLSLNLGLAFLLLFVL 261
OY 282 DRSPYILLTLTLVAVGICI-----NLIEVTKSOLKIQYPMNDYSEFMGNFSF----- 331
DB 262 AAGPTVYLLS-----AFGDNIGNYLGNLVRLSFKT-----YAYERHKKWPFESWTLYWAM 312
OY 332 ---WTGVSVLIMLFVGGNVIRKPGWLTGALVTPVMVLLTGIVFPALVIFRNQASGLVAM 388
DB 313 WCSWAPFVGGLFIARISKRTIREF--VFGVLLIPG--LFGVLMF-----TV 354
OY 369 FGTPPLML-AVVVGAIONILSKSTKVALFDSTKEMAYIPLDOEQV----- 433
DB 365 FGNIAIWLNDGVAGVLEKMTISSPETLLF--KFNRYLPBELNISVLSLVSIFFTVSA 411
OY 434 -----KGRKA-----IDVVAARFGSGGALLQOGLLVICSGIGAM 468
DB 412 DSGIYVNNITSRDGLSAPRWQAVMWGVMSAVAVLMSRSG-----LGNLQSM 461
OY 469 T-----PYLAVILLFIATWL-VSA-----TKLN-----KLFASALKEQ 503
DB 462 TLIVSLPFLALMLIMCPSLMKGLSADKKYFETRVNPTSVFTGCKMKERLVQIMSQTOEQ 521
OY 504 EV---AOEDSAPA 513
DB 522 DILKFLKQTASPA 534

RESULT 18
AAU91074
ID AAU91074 standard; Protein; 675 AA.
XX
XX AAU91074;
XX
XX 05-JUN-2002 (first entry)
XX
XX Neisseria cell surface polypeptide #12.
XX
XX Cell surface protein; antibacterial; antimicrobial.
XX
XX Neisseria meningitidis.
XX
XX WO200216612-A2.
XX
XX 28-FEB-2002.
XX
XX 21-AUG-2001; 2001WO-GB03759.
XX
XX 24-AUG-2000; 2000GB-0020952.
XX
XX (MICR-) MICROSCIENCE LTD.
XX
XX Lane JD, Hughes MJG, Santangelo JD;
XX

```

XX WPI: 2002-280941/32.
DR N-PSDB; AKS4090.
XX Novel peptide encoded by *Neisseria meningitidis*, useful for manufacture
PT of medicament for treatment or prevention of condition associated with
PT infection by *Neisseria* or Gram-negative bacteria -
XX Claim 5; Page 60-62; 79pp; English.
XX
XX The invention relates to polypeptides located on the cell surface of
CC *Neisseria meningitidis*, and the polynucleotides encoding them. The
CC sequences of the invention are useful for therapeutic or diagnostic use,
CC in the manufacture of a medicament for use in treatment or prevention of
CC a condition associated with infection by *Neisseria* or Gram-negative
CC bacteria. The sequences are also useful for screening potential
CC antimicrobial drugs or for detection of virulence. Sequences
CC AAU91063-AAU91079 represent *Neisseria meningitidis* polypeptides of the
CC invention.
XX
XX Sequence 675 AA;
SQ

Query Match 5.5%; Score 144; DB 23; Length 675;

Best Local Similarity 20.7%; Pred. No. 1.9e-06;
Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;

```

OY 32 LMFECITNTYVLRDTKDTLVAGP-----SGAEAIPIKPF-WLVVPCAIITM----- 79
DB 22 LFFVVCVAV-----VLVTVPDQVMWMDRAKEVITFEFSWFVLTFSIFGLFLLI 71
OY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFTVYIPLRDVLPHTPEF 125
DB 72 LSVSSLGNIRLGRDVEDPPEFGFLSWLMLFAAGVGVLMEFGVAEPLMHVPSDITATPE 131
OY 126 ADRLQALIPGGLGLVALIRNMTFAFYVLAELMGVSLMFMGFANEITKIHAKR-- 183
DB 132 HROQOALL-----HTVHHGVHAMS-----YGTIALALAYRFR--YKLPLALRSC 176
OY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVE--GVDPWGISLRLL 227
DB 177 FYPLKEKISGRFGDAIDIMALLATFFGIITTLGFGASQLGAGLQEIQWIAENSFSVQVL 236
OY 228 MAMTIVSGVLVMASTWYINKVLTDPFRYPNEBQKKGAK--PKNMKDSFL-----YL 281
DB 237 IIAVMSLAVVSA-----ISGVGKGVVLSLNLGLAFLLLFVL 276
OY 282 DRSPYILLTLTLVAVGICI-----NLIEVTKSOLKIQYPMNDYSEFMGNFSF----- 331
DB 277 AAGPTVYLLS-----AFGDNIGNYLGNLVRLSFKT-----YAYERHKKWPFESWTLYWAM 327
OY 332 ---WTGVSVLIMLFVGGNVIRKPGWLTGALVTPVMVLLTGIVFPALVIFRNQASGLVAM 388
DB 328 WCSWAPFVGGLFIARISKRTIREF--VFGVLLIPG--LFGVLMF-----TV 369
OY 369 FGTPPLML-AVVVGAIONILSKSTKVALFDSTKEMAYIPLDOEQV----- 433
DB 370 FGNIAIWLNDGVAGVLEKMTISSPETLLF--KFNRYLPBELNISVLSLVSIFFTVSA 426
OY 434 -----KGRKA-----IDVVAARFGSGGALLQOGLLVICSGIGAM 468
DB 427 DSGIYVNNITSRDGLSAPRWQAVMWGVMSAVAVLMSRSG-----LGNLQSM 476
OY 469 T-----PYLAVILLFIATWL-VSA-----TKLN-----KLFASALKEQ 503
DB 477 TLIVSLPFLALMLIMCPSLMKGLSADKKYFETRVNPTSVFTGCKMKERLVQIMSQTOEQ 536
OY 504 EV---AOEDSAPA 513
DB 537 DILKFLKQTASPA 549

RESULT 19
AAU74600

```


ID AAY74600 standard; Protein; 660 AA.
 AC AAY74600;
 DT 21-MAR-2000 (first entry)
 DE Neisseria gonorrhoeae ORF 163 protein sequence SEQ ID NO:674.
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.
 OS Neisseria gonorrhoeae.
 PN MO9957280-A2.
 PD 11-NOV-1999.
 PF 30-APR-1999; 99WO-US09346.
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253362.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 PS Claim 2; Page 460; 1453pp; English.
 CC AA253015 to AA254536, AA254537 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 SQ Sequence 660 AA;
 Query Match 5.1%; Score 134; DB 21; Length 660;
 Best Local Similarity 20.3%; Pred. No. 2.4e-05;
 Matches 125; Conservative 100; Mismatches 169; Indels 222; Gaps 35;
 QY 32 IMFPCITFNYVLRDKTLIVAGPQ-----SGAEAPFIKFLMLVPCALIFM----- 79
 DB 7 LFFVCLV-----VLVLTVPDQVGMWLDRAKEVIFTEFSWYVLTFSIFGLLI 56
 QY 80 LIYAKLSN-----LSKQALFYAVGTPF-LIFPALFPTVYIPLADVLAHPTTF 125
 DB 57 LSVSGIGNRIIGRDEVDPEFGFLSWLAMLFAAGMGVGLMFVVAEPALMHYFSDIVGAP 116
 QY 126 ADRLQALIPGLLGLVAILRNWTFAPAFVYLAEIMGSVMSLMPWGPANETITHEAKR-- 183

DB 117 HRQQQALL-----HTVFHMGVHAMSV-----YGTIALALAYFGFR---YTLPLALNSC 161
 QY 184 FYAL-----FGIGANI-SILASGRAIV-----WASKLASVSE-----GVDPMGIBLR 225
 DB 162 FVPLKKEKISGRFGAIDIMALLATFGIITTLGASQLGAGLQEMGMIAENSFGVYVL 221
 QY 226 LL---MAMTIYSGLVLMASYMWINKNVLTDRPFYPEEMOGKKGAKPRMNMKDSFL--- 279
 DB 222 IIAVMSLAIVSAT-----SGVGKGVK-VLSBINIGLAFLLF 258
 QY 280 -YLDSPYILLTLTLVAVGICI-----NLEVTWKSQKLYPMNDYSEFMGNFSF-- 331
 DB 259 FVLADPVTYVLLS-----AFGNIGNYLGNIIVRLSLKT-----YAYERHKKPFESMTVLY 309
 QY 332 -----WTGVSVYLIMLVGQNVIRKFGMLTGALVTPVMVLTGLVGFALVYFRQASGL 385
 DB 310 WAMMCSNAPFVGLFIRISKRTIRF--VFGVLLIPG--LFGVLMF----- 352
 QY 386 VAMGCTPLML-AVVVGAIONILSKTKYALFDSFKEMAYIPLDEQKV----- 433
 DB 353 -TVFGNTAIWLDGVAGMLERKMTSSPETLIF--KFTNYILPELTISTVSLVLSLFFV 408
 QY 434 -----KQKAA-----IDVVARFGKSGCALIQGLLVYCGSI 465
 DB 409 TSADSGIVVANNITSRDGLSAPRWQAVMWGMGSAVAVLMRSG-----LGNTL 458
 QY 466 GANT-----PYLAVILFIIAIWL-VSA-----TKLN-----KLFLAQSAL 500
 DB 459 QSMTLIVSLPFLMLIMCFSLMKGLSADKXYFETRVNPTSVFMTGKWKERTVIRMSOT 518
 QY 501 KQOEY---AOEDSAPA 513
 DB 519 QEDILKFLKHTASPA 534
 RESULT 20
 ABP78636
 ID ABP78636 standard; Protein; 675 AA.
 AC ABP78636;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 3802.
 KW Antibacterial; infection; vaccine; gene therapy.
 OS Neisseria gonorrhoeae.
 PN WO200279243-A2.
 PD 10-OCT-2002.
 PF 12-FEB-2002; 2002WO-1B02069.
 PR 12-FEB-2001; 2001GB-0003424.
 PA (CHIR-) CHIRON SPA.
 PA Fontana MR, Pizsa M, Masignani V, Monaci E;
 PI WPI; 2003-058415/05.
 DR N-PSDB; AB239606.
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection
 PS Disclosure; Page 463; 815pp; English.
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition

comprising the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection, this may be in the form of a vaccine or gene therapy.

Sequences given in records ABP76736-ABP81046 represent nucleic acid molecules of the invention.

Sequence 675 AA;

Query Match 5.1%; Score 134; DB 24; Length 675;

Best Local Similarity 20.3%; Pred. No. 2.5e-05; Indels 222; Gaps 35;

Matches 125; Conservative 100; Mismatches 169;

```

32 IMFPCITNYTVLRDTKDLIVGAGC-----SGAEALPIFIK-WLVVPCALIFM----- 79
Db LFFVGVLV-----VLVITVPDQVMMDRAKEVIFTEFSFYVLTSPFIIFGLI 71
Qy 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFTVIVPLRDVLPTEF 125
Db 72 LSVSGLGNIRIGRDEDEVEFGFLSWLMLFAAGMGVGLMFQVABPLMHFSIDIVGAPE 131
Qy 126 ADRLOAIIIPGGLIGVALIRWTFPAFYVLAELMSVWLSLMEFGFANEITKIHAKR-- 183
Db 132 HROOQALD-----HTVFHMGVHMSV---YGTIALALAYFGFR---YKLPLALRSC 176
Qy 184 FYAL-----FGIGANI-SILASGRAIV-----WASKLRASVE-----GVDPMGISLR 225
Db 177 FYPLIKKISGRFGDAIDIMALTFFGLITTLGSGAQGLGLOEMWIKENSQVYL 236
Qy 226 LL-----MAMTVISGLVLA SYWYMNKRVLTDPFVNPBEMQKKGAKPRMNMKDSFL--- 279
Db 237 IIAAVMSLAIVSAI-----SGVGKGVK-VLSELNIGLAFLLLF 273
Qy 280 YLDRSPITLLTLTLVAVGICI-----NLIEVTKWSQKIQYPMNDYSEPMGNEF-- 331
Db 274 FVLADPVPVLLS-----AFGDNIGNYLGNLVRLSLKT-----YAYERHKMFESWTLY 324
Qy 332 -----WTVGVSVLMLFVGVGVIRKFGMLGALVTPVMVLTGIVFALVIFRQASGL 385
Db 325 WAMMCSNAPFVGLPIARISKRTREF--VFGVLLIPG---LFGVLM----- 367
Qy 386 VAMFGTTPMLM-AVVVGAIONILSKSTYALPDSTKENAVYPLDQQRV----- 433
Db 368 -TVGNTAIIWINDGVAGMLEKMTSSPELTLF---KFPNYLPLPELTISIVSLVLSLFPV 423
Qy 434 -----KGKAA-----IDVNAARCKSGGALIQGLLVICSI 465
Db 424 TSADSGIYVNNITSRDGSLAPRQAVMVGVLMSAVAVLLMRSG-----LGNL 473
Qy 466 GAMT-----PYLAVILFIIAIWL-VSA-----TKLN-----KLFLAQSAL 500
Db 474 QSMFLIVSLPFLALMLIMCPSLMKGLSDKKYFETRVNVPVSFVWTGCKKRLVIRMSQT 533
Qy 501 KEQEV--AQEDSAPA 513
Db 534 QEDILKFLKHTASPA 549

```

RESULT 21

AAU34551 ID AAU34551 standard; Protein; 463 AA.

AAU34551;

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #132.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Escherichia coli.

WO200170955-A2.

```

XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR
XX 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
PI
PI Haeelbeck R, Ohlsen KU, Zyckind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
XX N-PSDB; AAS52410.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10144; sllp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are:
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 463 AA;
SQ

```

Query Match 5.0%; Score 132.5; DB 22; Length 463;

Best Local Similarity 22.3%; Pred. No. 2.1e-05;

Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;

```

Qy 69 WLVPKAIIFMLIVAKLSNI-LSKQALFYAVGTFPFLFALFPTVIVPLRDVLPTEPAD 127
Db 39 WISVPCLLAFVCMWLSAVAVNLPKGFNFTDQLFMLTALPVSGLALRV--PYSF-- 94
Qy 128 RLQAIIPGGLIGVAL--RNVTFPAFYVLAELMSVWLSLMEFGFANEITKIHAKRFY 185
Db 95 -----WVPIFGRRMT--AFST-----GILIPCVMLGFA-----VQDSTPY 130
Qy 186 ALFGIGANISILASGRAIVWASKLRASVS-----EGVDPMGISRLIMAM 230
Db 131 SVFII---ISLGFAGANFASNM-ANISFFPKQKOGGALGNGIGMNCVSTMOVLAP 186
Qy 231 TIVSGVLWMSYWMYMNKRVLTDPFVNPBEMQKKGAKPRMNMKDSFLYDRSPYLL- 289
Db 187 LVVS-LSIFAVF-----GSGQVVKPDGTE--LVLANASWIMVP 221
Qy 290 -LTLIVAYGICNLIVTWKSQLQYPMND-----YSEFMGNF-SFTTG--- 335
Db 222 FLAIFTIAWFGWMDL-ATSKASIKQLPVLKRGHLWIMSLVLTATGSGFISAGPAML 280
Qy 336 -----VSVLIMLVFG--GNVIRKFGW-----LTGALVTPV---MVLITGIVFALV 376

```

Db 281 SKTGPVDVQILQVAFGPFIGALARSAGALSDRLGTRVTLVNTLMAIFSGLLFLTP 340
 QY 377 IFRQASGLVAMFGTPTMLAVV-----GAIQNIISKSTIKYLFSDTKEMAYIPLDQE 430
 Db 341 T-DGGGSGFMAFFA---VFLLFLTAGLGGSGSTFQMIIVIRKULTMDRVKAEQ---GSD 392
 QY 431 QKVKGKALDVVAA-----RFKSGGALLIOGL---LVICGS-IGAMTPYLAVILLFPIA 481
 Db 393 ERARKERATDTAALGFIISAIGAGFPIPKAFGSSLLTQSPVGAMKVFLLFYIACVVI 452
 QY 482 IMLV 485
 Db 453 TWAV 456

RESULT 22
 AAU34568
 ID AAU34568 standard; Protein; 462 AA.

XX AC AAU34568;
 XX DT 14-FEB-2002 (first entry)
 XX DE E. coli cellular proliferation protein #149.
 XX KM Antisense; prokaryotic cellular proliferation protein;
 XX KW antibiotic; antibacterial; drug design.
 XX OS Escherichia coli.
 XX PN WO200170955-A2.
 XX PD 27-SEP-2001.
 XX PF 21-MAR-2001; 2001WO-US09180.
 XX PR 21-MAR-2000; 2000US-191078P.
 XX PR 23-MAY-2000; 2000US-206848P.
 XX PR 26-MAY-2000; 2000US-207727P.
 XX PR 23-OCT-2000; 2000US-242578P.
 XX PR 27-NOV-2000; 2000US-253625P.
 XX PR 22-DEC-2000; 2000US-257931P.
 XX PR 16-FEB-2001; 2001US-269308P.
 XX PA (BLIT-) BLITRA PHARM INC.
 XX PI Haeelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
 XX PI Yamamoto RT, Xu HH;
 XX DR WPI; 2001-611495/70.
 XX DR N-PSDB; AAS52427.
 XX PT New polynucleotides for the identification and development of
 XX PT antibiotics, comprise sequences of antisense nucleic acids -
 XX PS Example 3; Seq ID No 10161; 51pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 462 AA;

Query Match 4.9%; Score 129; DB 22; Length 462;
 Best local similarity 22.3%; Pred. No. 5,16-05;
 Matches 108; Conservative 71; Mismatches 171; Indels 134; Gaps 26;

QY 69 WLVPICAIIFMLIYAKLSNI---LSKQALFYAVGTPFIIFALPPTV-----IY 114
 Db 37 WISVSCILLAFQVMMLFSATVNLNKIGFNTTQTLT--TALPSVSGALLRPYSRMV 94
 QY 115 PLRDVLHTEFADRLQAILPPGLIGLVAILLNMTFAAPYVLAELWGSVMSLMFWG--FA 172
 Db 95 PIFGRRRTVTFSTAI-LIIPCVMLGIAVQNPNTFFGIFIVYALCG-----FAGANFA 146
 QY 173 NEITKIH-----EAKRFYALRGIGANISLASGRAIWAASKURASVSEGVDPWGISLRLLM 228
 Db 147 SSMGNISFFPPPKAKQSGAL-GINGGL-----GNLGSVMQULVAPLVFVPEFA 193
 QY 229 AMTI-----VSGVLMASYW-----WINKVLTDPFRFYNPEMOKKKGKA 268
 Db 194 FLGVNGVPQADGSVMSLANNAAMIVPILATATIANAGANDI-----A 236
 QY 269 KPRKMKDSFLYIDRSPYILLTLVIA-YGICINLIEVTKWSQKLOYPNN-----DY 322
 Db 237 SSRASIAIDQLPVLQRL-HLWLLSLVLYATFGSFIG-PSAGPAMLAQOPDVNILLRLAF 294
 QY 323 SEFWGNFSFWTGVSVLIMLFVGVGNVIRKGMULTGALVTPVMULT-----TG-- 369
 Db 295 GPFITG-----AIAHSVGAISDKFGKGVVTLINFIEMAFALLFLTPGTGSG 343
 QY 370 --IYFPALVIFRNQASGLVAMFGTPTMLAVVVAIGNIISKSTIKY--ALFSDTKEMAYI 425
 Db 344 NFIAFYAVPMQLFITAGIGS--GSTFQMIAVIFQI-----TIYKVKMGSSDEQAH- 393
 QY 426 PLDQQRVKKKAIDVVAARFGKSGALLIOGL---LVICGS-IGAMTPYLAVILLFPIA 481
 Db 394 ---KEAVTEPTAALGFIISA-IGAVGPFPIQAFGMSLMTMGSPVGAMKVFLLFYIACVIL 449
 QY 482 IMLV 485
 Db 450 TWLV 453

RESULT 23
 ABJ19074
 ID ABJ19074 standard; Protein; 435 AA.
 XX AC ABJ19074;
 XX DT 06-MAR-2003 (first entry)
 XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 351.
 XX DE Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
 XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
 KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
 KW autoimmune disease; HIV; hepatitis.
 XX OS Staphylococcus sp.
 XX PN WO200259148-A2.
 XX PD 01-AUG-2002.
 XX PF 21-JAN-2002; 2002WO-EP00546.
 XX PR 26-JAN-2001; 2001AT-0000130.

PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Meinke A, Nagy E, Von Ahesen U, Klade C, Henics T, Zauner W,
 PI Minn DB, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M,
 PI Tempelmeier B;
 XX
 DR WPI, 2003-075410/07.
 XX
 PT Identifying, isolating and producing hyperimmune serum-reactive
 PT antigens from a pathogen, for preparing vaccine or medicament for
 PT treating or preventing e.g. staphylococcal infections, comprises
 PT providing antibody preparation -
 XX
 PS Example 7; Page 214; 252pp; English.
 XX
 CC The invention relates to a novel method for identifying, isolating and
 CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
 CC allergen, a tissue or host prone to auto-immunity, where the antigens
 CC are used in a vaccine, comprises providing antibody preparation from a
 CC plasma pool of a type of animal, or individual sera with antibodies
 CC against the specific pathogen, tumour, allergen, tissue or host prone to
 CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
 CC the 62 sequences of 53-2261 amino acids fully defined in the
 CC specification, or their hyperimmune fragments are useful in the
 CC manufacture of a pharmaceutical preparation, particularly a vaccine
 CC against staphylococcal infections or colonisation against S. aureus or S.
 CC epidermidis. The preparation of antibodies is useful for the manufacture
 CC of a medicament for treating or preventing staphylococcal infections or
 CC colonisation against S. aureus or S. epidermidis. The antibody
 CC preparations may also be used for diagnostic and imaging purposes. Other
 CC conditions that can be treated include cancer, autoimmune diseases or
 CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
 CC protozoan pathogens. This sequence represents a staphylococcal protein
 CC relating to the method for identifying and producing pathogen specific
 CC antigens of the invention.
 CC
 XX
 SQ Sequence 435 AA;
 Query Match 4.8%; Score 125.5; DB 24; Length 435;
 Best Local Similarity 19.9%; Pred. No. 0.00011;
 Matches 96; Conservative 82; Mismatches 162; Indels 143; Gaps 19;
 QY 9 FGKLSFMPFIHTHKLKVLPMFLMFCITFNYTVLARTKOTLIYGAFSGAEI----- 63
 DB 17 FGAGULIRPPMIGHAGQNMWGMGFALT---GILLPPIYIVAFDEGVESGNRIH 73
 QY 64 PFIKFWLVPCALIFMLIYA-----KLSNISKQALFYAVGT-----PFLIPFAL 108
 DB 74 PMFGFIFAV---VIYMSIGAFYGIIPRAANVA-----YEIGRHILPVHNGWTLIIFAI 124
 QY 109 PPTVIVYPLADVLHPEFADRLCALIPRGILGVALLRWMTPAFV----- 153
 DB 125 FFALVYWIS--LNPSEKIVDNIGKLLTPLLMLVALL---SLAVINPESALSAPROKXIT 179
 QY 154 -----VLAELMGSV-----MLSLMFGFANEIKIHEAKRF 184
 DB 180 HPIFISLEGFTYTMVLVALLAFSVIVNGYKFKGLTDMKILKYCFSGLIALLIGMIV 239
 QY 185 YALFEGIANIS--ILASGRAIWAASKRASVSEGVDPWGISILRLMAMTIVSGVLWASY 242
 DB 240 FALAVAGASTAAGNFKDGTDILTYSNLRFGSFGMLVFGMTV--ILACLTTCIGLV--NACA 297
 QY 243 WMINNVNLTDRFRVNEPMOKKGAAPRANK-----DSFYLDPSYIIL--LT 291
 DB 298 TFKKHV-----PKFSKIPALFISIIIGFLTLTGEMILKIVP 337
 QY 292 LLIYVAGICINLIEYTKMSQLQYPMNNDYSEFPMGTVGVSVLIMFVGANVRX 351
 DB 338 LITLIPYSIALVLISF-----ANMSTFFPSAAYRATYITITLISL--QIINS 385
 QY 352 FGMLTGALVTPMVLITGIVFPAVIFRNOASGLVAMEGTTPLMLAVVGAIONILSKST 411
 DB 386 FNLHGVILKSFMMLELADIDLAMLV-----PFMLFALIGFIIDVFIRP 430

QY 412 KYA 414
 DB 431 KOA 433
 RESULT 24
 AAY58582
 ID AAY58582 standard; Protein: 713 AA.
 AC AAY58582;
 DT 10-APR-2000 (first entry)
 DE Sorangium cellulosum protein Orf 3.
 XX
 XX Orf 3; Na/H antiporter; epothilone biosynthesis; export;
 KM taxol substitute; anticancer.
 XX
 OS Sorangium cellulosum.
 XX
 FM W09966028-A2.
 XX
 PD 23-DEC-1999.
 XX
 PF 16-JUN-1999; 99WO-EP04171.
 XX
 PR 18-JUN-1998; 98US-0099504.
 PR 24-SEP-1998; 98US-0101631.
 PR 05-FEB-1999; 99US-0118906.
 XX
 PA (NOVS) NOVARTIS AG.
 PA (NOVS.) NOVARTIS-ERRINDUNGEN VERN GES MBH.
 PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
 DR WPI; 2000-097741/08.
 DR N-PSDB; AAZ55887.
 PT New isolated epothilone synthase genes, used for the recombinant
 PT production of epothilone for use in cancer therapy -
 XX
 PS Claim 12; Page 162-164; 174pp; English.
 CC This sequence represents a protein, Orf 3, encoded by open
 CC reading frame 3 (ORF3) of a Sorangium cellulosum 68.75 kb contig
 CC which covers the epothilone biosynthesis genes. This sequence has
 CC homology to Na/H antiporters of various bacteria, and may take
 CC part in the export of epothilones from the producer cell. The
 CC invention relates to nucleic acid sequences encoding epothilone
 CC biosynthetic enzymes from Sorangium cellulosum. Epothilones A and B are
 CC 16-membered macrocyclic polyketides with an acylcysteine-derived
 CC starter unit; polyketides being synthesised from two-carbon building
 CC blocks, the beta-carbon of which always carries a keto group. Each round
 CC of the two-carbon addition is carried out by a complex of enzymes known as
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
 CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
 CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
 CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
 CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
 CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
 CC involved in transport. Epothilones mimic the biological activity of
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic
 CC compositions. Epothilones exhibit a much lower drop in potency against a
 CC multiply drug-resistant cell line compared with taxol, and are
 CC considerably less efficiently exported from such cells by the multidrug
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
 CC epothilones as anticancer agents, they are problematical to produce on a
 CC large scale. Epothilones are too complex for industrial scale chemical
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
 CC poor yields of epothilones. The nucleic acids of the invention may be
 CC used for the recombinant production of epothilones in a heterologous host
 CC that is more amenable to fermentation. Sequences AAY58580,

Qy	489	KLWK	492	
Db	385	RLDK	388	
RESULT 26				
ID	AAU36578	standard; Protein; 425	AA.	
AC	AAU36578;			
XX	14-FEB-2002	(first entry)		
DE	Staphylococcus aureus	cellular proliferation protein #748.		
KM	Antisense; prokaryotic cellular proliferation protein;			
XX	antibiotic; antibacterial; drug design.			
OS	Staphylococcus aureus.			
PN	WO200170955-A2.			
XX	27-SEP-2001.			
PF	21-MAR-2001; 2001WO-US09180.			
PR	21-MAR-2000; 2000US-191078P.			
PR	23-MAY-2000; 2000US-306848P.			
PR	26-MAY-2000; 2000US-207727P.			
PR	23-OCT-2000; 2000US-242578P.			
PR	27-NOV-2000; 2000US-253625P.			
PR	22-DEC-2000; 2000US-257931P.			
PR	16-FEB-2001; 2001US-269308P.			
PA	(ELIT-) ELITRA PHARM INC.			
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;			
PI	Yamamoto RT, Xu HH;			
DR	WPI: 2001-611495/70.			
DR	N-PSDB; AAS54437.			
PT	New polynucleotides for the identification and development of			
PR	antibiotics, comprise sequences of antisense nucleic acids -			
XX	Example 3; Seq ID NO 12171, 511DP; English.			
PS				
XX				
XX	The invention relates to antisense inhibitors of genes essential to			
CC	prokaryotic cellular proliferation, their use in identifying the			
CC	genes themselves and the encoded proteins. The prokaryotes used are			
CC	Bacteriophage coli, Staphylococcus aureus, Salmonella typhi, Klebsiella			
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The			
CC	invention is also useful for the identification of potential new targets			
CC	for antibiotic development. The antisense nucleic acids can also be used			
CC	to identify proteins used in proliferation, to express these proteins,			
CC	and to obtain antibiotics capable of binding to the expressed proteins.			
CC	The proteins can be used to screen compounds in rational drug discovery			
CC	programmes. The antisense nucleic acid sequence is also useful to screen			
CC	for homologous nucleic acids which are required for cell proliferation in			
CC	a wide variety of organisms. The present sequence represents an			
CC	essential prokaryotic cellular proliferation protein.			
CC	Note: The sequence data for this patent did not form part			
CC	of the printed specification, but was obtained in electronic			
CC	format directly from WIPO at			
CC	ftp.wipo.int/pub/published_pct_sequences.			
XX				
XX	Sequence 425 AA;			
Qy	Query Match	4.6%;	Score 120.5;	DB 22; Length 425;
	Best Local Similarity	19.5%;	Pred No. 0.0004;	

Matches	94;	Conservative	86;	Mismatches	160;	Indels	143;	Gaps	19;
QY		9	FGKLRSFLMP	HTHELKKVLEPMFLMFCIT	NTNYVLADTKTLLVVGAPGSAEAI	-----	63		
Db		7	FGAGNL	IFPPMLGHTAGGNMMWIGM	GFALT	---	61	PFITVIVAFYDEGVESGNRIH	63
QY		64	PFIKFWLV	PCPAIFEMLLYA	-----	KLSNLSKQALFYAVGT	-----	PFILFPAL	108
Db		64	PMFGGFI	PAV	---	VIYMSIGAFYGGIPRAANVA	-----	YEIGTRILPLVHNQWTLIFAAV	114
QY		109	FPETVYPL	RDVLAHPTEFADRLQAL	LPGLGLVALILRNMTFAFY	-----	153		
Db		115	FFAIVYVIS	--LNSKIVDNLGKULLTP	LLIMVALL	---	STAVIFNPESALSAEKDYIT	169	
QY		154	-----	-----	-----	-----	-----	-----	-----
Db		170	HPFISGSL	EGYFTMDLVAAALAFSV	VIIVGKYFKGLTDRTKLTKYVCSGFI	AAILGLM	YIT	229	
QY		185	YALGIGANIS	--LLASGR	AIIVMSKCLRASVSEGVDDPMGISTRLMM	MTIVSGVLMA	SY	242	
Db		230	FALYVVG	ASTAPGNPKDGTDL	TYNSLRKFGSGFNLVFGMTV	--ILACIT	TIGTLV	NACA	287
QY		243	WMINKNV	LTDRFRFYNPEEMQKGGAKK	AKNMK	-----	DSFLYLD	RSPYILL	291
Db		288	TFTKKNV	-----	-----	-----	-----	-----	-----
QY		292	LIVAIY	GCINLIEVTWKSQ	LKQYPMNDYSEFMGNPSEWTVGSV	LVLM	FGGNVIRK	351	
Db		328	LLTLIY	PSIALVLSIF	-----	AMFSTPRFSMAYRLATVIT	LLISIL	--OILNS	375
QY		352	FGWLTG	ALVTPVWMLTGIV	FPALVIFPNQASGVAMFGT	PLMLAVVGAIONIL	KST	411	
Db		376	FNLH	GVILKWFMLPLADID	LAWLV	-----	PFILALIGFI	IDVIRRP	420
QY		412	KYA	414					
Db		421	KQA	423					
RESULT 27									
ABJ19068									
ID	ABJ19068	standard; Protein; 472 AA.							
XX	ABJ19068;								
XX	DT								
XX	06-MAR-2003	(first entry)							
XX	Pathogen specific	antigen related staphylococcal protein SEQ ID NO 345.							
XX	Antibacterial; virucide;	fungicide; protozoacide; cytostatic; anti-HIV;							
XX	hyperimmune; serum-reactive;	antigen; pathogen; tumour; allergen;							
XX	auto-immunity; vaccine;	staphylococcal infection; antibody; cancer;							
XX	autoimmune disease; HIV;	hepatitis.							
OS	Staphylococcus sp.								
PN	WO200259148-A2.								
XX	01-AUG-2002.								
PD	21-JAN-2002; 2002WO-EP00546.								
PF	26-JAN-2001; 2001AT-0000130.								
PR									
XX	(CIST-) CISTEM BIOTECHNOLOGIES GMBH.								
PA	Weinke A, Nagy E, Von Ahnen U, Krlade C, Henics T, Zauner W,								
PI	Mith DB, Vayevyska O, Ecz H, Dryla A, Welchart T, Hatner M,								
PI	Tempelmaier B;								
XX	WPI; 2003-075410/07.								
DR									
XX	Identifying, isolating and producing hyperimmune serum-reactive								

PT antigens from a pathogen, for preparing vaccine or medicament for
 PT treating or preventing e.g. staphylococcal infections, comprises
 PT providing antibody preparation -

PS Example 7, Page 214, 252pp; English.

CC The invention relates to a novel method for identifying, isolating and
 CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
 CC allergen, a tissue or host prone to auto-immunity, where the antigens
 CC are used in a vaccine, comprises providing antibody preparation from a
 CC plasma pool of a type of animal, or individual sera with antibodies
 CC against the specific pathogen, tumour, allergen, tissue or host prone to
 CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
 CC the 62 sequences of 53-2261 amino acids fully defined in the
 CC specification, or their hyperimmune fragments are useful for the
 CC manufacture of a pharmaceutical preparation, particularly a vaccine
 CC against staphylococcal infections or colonisation against *S. aureus* or *S.*
 CC epidermidis. The preparation of antibodies is useful for the manufacture
 CC of a medicament for treating or preventing staphylococcal infections or
 CC colonisation against *S. aureus* or *S. epidermidis*. The antibody
 CC preparations may also be used for diagnostic and imaging purposes. Other
 CC conditions that can be treated include cancer, autoimmune diseases or
 CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
 CC protozoan pathogens. This sequence represents a staphylococcal protein
 CC relating to the method for identifying and producing pathogen specific
 CC antigens of the invention.

XX Sequence 472 AA;

Query Match 4.5%; Score 119.5; DB 24; Length 472;

Best Local Similarity 19.4%; Pred. No. 0.00059;
 Matches 58; Conservative 61; Mismatches 111; Indels 69; Gaps 12;

QY 151 AFVTLAEWGSVMSLFWGPFANEITKHEAKRYALFG--IGANISLISAGRAIWASK 208
 DB 175 AFLVTFEFGNLTIAAF-----LTAAAGNPLAQNLTASSTNVHITWNNW 219
 QY 209 LRASVSEGVDPWGSIRLIMAMTIVSGVLMAVYWKIKNVLTDFRFPNPEMOKGKKGA 268
 DB 220 FLALVLP-----LVSLIVPFTI---KYIYPVKEPT--NAKSMAMENLAT 262
 QY 269 KPKNMKDSFLYLDPSYILLTLTLVAVGICINLIEVTWKSQKLQYPMNDYSEPMGN 328
 DB 263 MGKIALAEKMTI---GIFVALTITMIVGSPFHIDTTLTAFTALALLLTGVLTMODITNE 319
 QY 329 FSPFTGVV--SVLIMLFGVGNVIRKFG--WLTGALVTP-----VMVLLTGI VEPAL 375
 DB 320 TGAWNTLVWFSVLVLM--ADQLNLTGFIPLWLSKSIATSLGSLGMPVLVITLILFEYFYSH 376
 QY 376 VIFPNQASGLVAMF-----GTPPLMAVVVGAIONILSKSTKA-----LFPSS 418
 DB 377 YLFASSTAHISAMYAALLGVALAAGAPPLFSALMLGFGNLLASTTHYSSGPAPILFSS 435

RESULT 28

ID ABP40810 standard; Proteoin; 800 AA.

XX AC ABP40810;

XX DT 24-JUL-2002 (first entry)

XX DB Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5655.

XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX KM antibacterial; gene therapy.

XX OS Staphylococcus epidermidis.

XX US US6380370-B1.

XX PD 30-APR-2002.

XX

PF 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; AEN93355.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -

PS Disclosure, SEQ ID 5655, 267pp; English.

CC AEN90538 to AEN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP5124 to ABP37960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life
 CC cycle or inhibit *S. epidermidis* infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX Sequence 800 AA;

Query Match 4.5%; Score 119.5; DB 23; Length 800;
 Best Local Similarity 20.3%; Pred. No. 0.0013;
 Matches 108; Conservative 86; Mismatches 148; Indels 191; Gaps 30;

QY 2 TKTEKEPFGKLRSLW-----PI-----HTHEKKVLPFIMF-----F 35
 DB 222 TKSAQFPF-----HIMLPKMAAPTPVSAYIHSATWKA-GIFLFRPTPLGLSNVIY 275
 QY 36 CTFP-----NYTVLR-DTKDTL-----IVGAPSGAE-AIPFIKFW 69
 DB 276 TVTFVGLITMLFGSLTALROYDLKGLIAYSTISQIMTMWVGAGGYAQTSDLSKEY 335
 QY 70 LVVFCALIIIPMLIYAKLSNLSKQALFYAV----- 98
 DB 336 ILVLFAGLFLH---NMHAFKCALFMGVGIIIDHESGTRDIRLNGKRYEPPKHIVML 391
 QY 99 -----GTPFLIFPALPPTVYPRADVLPHTPEFADRLQALIPGL--LGLVALIDNMTFA 150
 DB 392 AALSMAGVPFLNGLSKEMFL---DSLTKANELDQGFVITPVIISIGVASLITFTYA 447
 QY 151 AFVTLAEWGSVMSLFWGPFAN---EITKHEAKRP-----VALFGIGA 192
 DB 448 LY-----MIKETFGVNYNIEKFKRKQIHBPMLFSLPAVITMELIIPVIFPVNFG 497
 QY 193 NISLLASGRAIWAASKLRASYS---EGVDPK-GISLALLMA-KTIVSGVLVMASTW--- 244
 DB 498 NFVILPAPTRSV---SGIAGEVDAPVPHISQHWGNLPLILISIVIIIGLIALLVNMRKV 554
 QY 245 -----INKNVLTDP--PRFPNPEMOKGKAKPKKNMKDSFLYLDPSYILLTLTL---V 294
 DB 555 THQIITKASITDGRKRYREFELYS-A-RGIRALMNANKN-----YIMTITLFIPIAI 605
 QY 295 IAVGICINLIEVTWKSQKLQYPMNDYSEPMGNFSFWTVGSVLMLFVGANVIRKFGW 354
 DB 606 VVYG-----YLVGFPFHQTL--HISFGLPELVILSV-VLLIIGISLIFROR 650
 QY 355 LTGALVTPWVYLLTGLVFPALVIFRNQASGLVAMFGTTPPLMAVVVGAIONIL 407
 DB 651 LT-----NVVLNMGIMGFAVTV-----FIAMKAPDLALTOLVETITTL 690

RESULT 29

AA49633
ID AA49633 standard; Protein; 517 AA.
XX
AC AA49633;
XX
DT 14-JAN-2000 (first entry)
XX
DE Wheat hexose carrier protein.
XX
KW Hexose carrier protein; corn; rice; sorghum; soybean; wheat;
KW carbohydrate transport; plant carbon partitioning; manipulation;
KW carbohydrate distribution.
XX
OS Triticum aestivum.
XX
PN MO9953082-A2.
XX
PD 21-OCT-1999.
XX
PE 07-APR-1999; 99WO-US07561.
XX
PR 09-APR-1998; 98US-0081131.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Lightner JE, Rafalski JA, Thorpe CJ;
XX
DR WPI; 1999-620438/53.
XX
DR N-PSDB; AA232203.
XX
PT New hexose carrier proteins used to manipulate carbohydrate transport
XX
PS Claim 5; Page 49-51; 60pp; English.
XX
CC The present sequence represents a hexose carrier protein from the
CC invention, which describes hexose carrier proteins isolated from sorghum,
CC rice, wheat, soybean and corn. Also describes are: (1) a chimeric gene
CC comprising a hexose carrier protein polynucleotide operably linked to
CC regulatory sequences; (2) a transformed host cell comprising the chimeric
CC gene; and (3) altering the level of expression of a hexose carrier
CC protein in a host cell, comprising transforming a host cell with the
CC chimeric gene and growing the cell under expression conditions. Hexose
CC carrier proteins may be used to manipulate carbohydrate transport and to
CC alter whole plant carbon partitioning or to manipulate carbohydrate
CC distribution between cellular compartments.
XX
SQ Sequence 517 AA;
XX
Query Match 4.5%; Score 119; DB 20; Length 517;
Best Local Similarity 18.6%; Pred. No. 0.00077;
Matches 106; Conservative 82; Mismatches 173; Indels 208; Gaps 27;
XX
QY 28 LPMFLMFCIT-----FNYTVLRDTKDTLIVGAPSGAELPIKFWLVVPCALIML 80
XX
DB 19 LTIFFVFTCVAAATGGLFGVD-----IGISGVTSMNPFLKFE-----PPE 60
XX
QY 81 IYAK-----LSNLSKQALFYAVGTPFLFPALFPVIVYPLRDVHPT 123
XX
DB 61 VYDKKMGKSGASQYCKYDNQLQFTSSLYLAALVSS---FFA--ATV-----T 104
XX
QY 124 EFADRLQALPPGGLGLVAILRNWFAFYVLAELMGSVLMWGMWGANETKTHEAK- 182
XX
DB 105 RVVGKRWMEFTGGTLFLLGALNG--AAENIMLTVGRLLGVGV--GFANOSVVPVYLSM 161
XX
QY 183 ---RFAVAFGIG---ANISLASGRAIYVASKLRAVSEGVDPGMSILMLMAMTIVSG 235
XX
DB 162 APARLRGMLNIGFQMLITIGILAALINVDNKKIRAGG-----WRISLAI---AAVPA 213
XX
QY 236 LVLAASYWINKNVLTDPRFYNPEE---MQGKKGAKPKM-----NMKDSFLYLD 284
XX
DB 214 IITLGS-----FFLPDPNPSLIERGHDEAARMMLNRIRGSDVDISEYAD-- 258

QY 285 FYILLTLVLAAYGICINLIEVTWKSQKLOYPNNNDYSEFMGNFSFWTGVSVLIMLFV 344
XX
DB 259 -----LVVASEESKL-VQHP-----WRNLIQ----- 278
XX
QY 345 GGNVIRKE-GMLTGLVTPVWVLLTG---IVFALVIFRNGASGLVANFGTTPMLAAVV 400
XX
DB 279 -----RKVRPOLTWMIMPFQQLTGINIVIFVAPVLFET-----LGRKGDASISAVIT 328
XX
QY 401 GAIQNLISKRTKVALPDSKEMAVIPLDQEKVKGKALDVVAARFGSGGALLQGG--- 457
XX
DB 329 GLVAVFATLVSVFTVDLRGRKRLFLQGGTQMLLSQLVVGTLIAVKFGTSGVGEWPKGYAA 388
XX
QY 458 --LLVIC-----GSIGAMP-----YLAVI 475
XX
DB 389 AVVLFICLYVAGFAMSGPLGMLVPSSEIFPLEIRPAGSINVSVMLETFVIAQAFILML 448
XX
QY 476 -----LLEFIIALWVSATKINKLFLAQS 498
XX
DB 449 CHMKFGLFFYPFAGWVIMTVFIALFLPET 477
XX
RESULT 30
AAU38238
ID AAU38238 standard; Protein; 466 AA.
XX
AC AAU38238;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #129.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haeselbeck R, Ohlsen KU, Zyckind JW, Wall D, Trawick JD, Carr GU;
PI Yamamoto RT, Xu HH;
XX
DB WPI; 2001-611495/70.
XX
DR N-PSDB; AAS56097.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 13831; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 466 AA:

Query Match 4.5%; Score 118; DB 22; Length 466;
 Best Local Similarity 20.4%; Pred. No. 0.00086;
 Matches 99; Conservative 70; Mismatches 179; Indels 138; Gaps 23;

QY 69 WLVPFCALIFMLYAKLSNI--LSKQALFYAVGTPLIFPALPPTVIYPLADVLHPTPEF 125
 DB 41 WISVSCILAFVWMLFSAVAVNLNKIGPDTTDLFL-----TAL 82.
 QY 126 ADRLQALIPGLIGLVALI--RNWTFAPFYLAELMSVMSLMEWGA--NEITKIHBA 181
 DB 83 PSLGAILRVPSFVWPLFGGRKMTVLSTVLI-----IIPCAMLGFAVQNPATP--- 131
 QY 182 KRFPALFGIGANISHLASGRAIWMASKL-----RASVSGVDPWGISLRL 227
 DB 132 -----FVFTLIALCGFAGANFASWGNISFPFKARQSGALCINGLGVSWQL 185
 QY 228 MAMTIV-----SGVLIMASYW-WIN-KWVLTDPREFNEEMQKKGK 267
 DB 186 IAPLVIPIPIFPLGVRGPQDGLLALTNAMTWVLLAVATLAAMFGMNDIGSSKAS 245
 QY 268 AKPKNMKDSFLYDRSPYIILLTLVLA-YGICINLIEVTWSQKLQYPMN-----D 321
 DB 246 VA-----SOLPVKRLHMLSLTLATGSSFIG-PSAGFAMLAKTQFPVNILQAF 297
 QY 322 YSEFMGNFSFMTGVSVLIMLFGVGNVIRKGMVLGALVTPV-WVLNGIYFPLVIRN 380
 DB 298 FGPIFG-----ALARSAGVVISDKRGVAVTLINIFMLFTALFLTL---PG 343
 QY 381 QASGLVAMF-----GTTPLMLAVVGAIONILSKSTKY--ALFDSKEMA 423
 DB 344 SGAGSFSAFLYVEMGLFLTAGSGSTFOMIAVIRQI-----TLNVKLRGSSDQA 396
 QY 424 YIPDQEQVKGAIDVVAARFGKSGGALIQQGL--LVICGS-IGAMTPYLAVILLFI 479
 DB 397 -----QREAVTDTAALGFISALGAVGGFIPKARGTSLALGSPVGMKIFLFIYACV 451
 QY 480 IAIWLV 485
 DB 452 LITWLV 457

RESULT 31

ABB93711 standard; Protein; 694 AA.

XX ID ABB93711;
 XX AC ABB93711;
 XX DT 31-MAY-2002 (first entry)
 XX DB Herbicidally active polypeptide SEQ ID NO 2922.
 XX KW Herbicidal; plant; agriculture; herbicide.
 XX OS Arabidopsis thaliana.
 XX PN WO200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidner M;

DR WPI, 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compound,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

PS Claim 5; SEQ ID NO 2922; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

CC Sequence 694 AA:

Query Match 4.5%; Score 118; DB 23; Length 694;
 Best Local Similarity 19.9%; Pred. No. 0.0015;
 Matches 97; Conservative 61; Mismatches 161; Indels 168; Gaps 19;

QY 92 QALFYAVGTPLIFPALFPYIYPLR---DVLPTEPADLQALIPGLIGLVAILRNW 147
 DB 35 QGVYAVILGGAFFAVFTSFLWLEKRYVGARTSEWFTAGNNVKTGLIASV-IYSQW 93
 QY 148 TEPA-----FVYLAELW---GSVMSLME-----W 169
 DB 94 TWALPTILOSNNVAMQYGVGPFMYASGATITVLFGVMAIRKRAPNAHTVCEIVKRW 153
 QY 170 GFANEITKIHBAKRFYALFGIGANI-----SLASGRAIWMASKLRASVSEGVDPWGISLR 225
 DB 154 GTATHIV-----FLVFCLATNVVTYAMLLOGSNAV-----NALTGVNLYAASF 198
 QY 226 L---LMAMTIVSGVLMASTWINKNVLTDRFYNPBEEMQKKGAKPKNMKDSFLYLD 282
 DB 199 IPLGVVVYTLAGL-----KATFLASY 220
 QY 283 RSPYIILLTLIVAVGICINLIEVTWS-----QKLOYPMNDYSEFMGNFSFMTGV-- 336
 DB 221 VHSYIVHVALVFPVLVYTSKELGSPSVVYDRDKMDVAKRSCTEPLSHHQAQCPVDG 280
 QY 337 ---SVLIMLFGVGG---NVIRKFG---WLTGALVTPVMV---LTGIVFEFA 374
 DB 281 NFRGSYILTMSSGGAIVGLINIVGNFPTVDNQYVWYSALARSSTHKGVLGLGVWFA 340
 QY 375 LVIRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALFDSKEMAYIPLDQEQVK 434
 DB 341 V-----PFSLATSLGALALD-----LPISGDEADR 367
 QY 435 GKALIDVVAARFGKSGGALIQQGLVITCGSIGAMTPYLAVILLFTIAMIWLSATKNTLF 494
 DB 368 GLVPPATAIAMLGKSGSLILLTMLFMVNTSAGS-SELIAVSLPTDYI---RTYINPRA 423
 QY 495 LAQSALK 501
 DB 424 TGRQILK 430

RESULT 32

AAU34182 standard; Protein; 430 AA.

XX ID AAU34182;
 XX AC AAU34182;
 XX


```

Db      241 TM---LIVGVVIALFW-RQVINDPME-----LHVRYV 274
Qy      287 ILLTLTLVAYGICINLEVTWKSQKLOYPNNDYSEPMGFVTVSVILMFEVG 346
Db      275 FSLSVLT---GSIVTMAMGAIVPLVYIQTRGSALQSGLLPLGAIIMGIMSPING 330
Qy      347 NVIRKFG-WLTGALVTPVWVLLTGVPFALVIFRQASGLVAMECT--TPMLANVGA 402
Db      331 IIFDKIGARWL---ITGVTLITGTIPF-----MFLTMDTPLYVVF-- 371
Qy      403 IONILSKSTYKALFDSTKEMAVYPLDQEKVKGAIDVVAARFGSGALIQOGLVTC 462
Db      372 -----YAVRFGISMAMPV---STAGMNLPHMLNHSGAVNNTIRQ-----IA 413
Qy      463 GSIGAMPYLAIVLLFTIAIMVLSATKLN--KLFLAQSALKEQVAGSDS 510
Db      414 GSIG-----TAVLITV---LTNVTKDNMPGKALMATDPASFAQAKQDAS 454

RESULT 35
AAU38247
ID      AAU38247 standard; Protein; 477 AA.
XX
AC      AAU38247;
XX
DT      14-FEB-2002 (first entry)
XX
DE      Salmonella typhi cellular proliferation protein #138.
XX
KW      Antisense; prokaryotic cellular proliferation protein;
XX      antibiotic; antibacterial; drug design.
XX
OS      Salmonella typhi.
XX
PN      WO200170955-A2.
XX
PD      27-SEP-2001.
XX
PF      21-MAR-2001; 2001WO-US09180.
XX
PR      21-MAR-2000; 2000US-191078P.
XX      23-MAY-2000; 2000US-206848P.
XX      26-MAY-2000; 2000US-207727P.
XX      23-OCT-2000; 2000US-242578P.
XX      27-NOV-2000; 2000US-253625P.
XX      22-DEC-2000; 2000US-257931P.
XX      16-FEB-2001; 2001US-269308P.
XX
PA      (ELIT-) ELITRA PHARM INC.
XX
PI      Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU,
PI      Yamamoto RT, Xu HH;
XX
DR      MPI: 2001-611495/70.
XX      N-PSDB; AAS56106.
XX
PT      New polynucleotides for the identification and development of
PT      antibiotics, comprise sequences of antisense nucleic acids -
PS      Example 3, Seq ID No 13840; 511pp; English.
XX
XX      The invention relates to antisense inhibitors of genes essential to
XX      prokaryotic cellular proliferation, their use in identifying the
XX      genes, their use in the discovery of novel antibiotics, the essential
XX      genes themselves and the encoded proteins. The prokaryotes used are
XX      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX      invention is also useful for the identification of potential new targets
XX      for antibiotic development. The antisense nucleic acids can also be used
XX      to identify proteins used in proliferation, to express these proteins,
XX      and to obtain antibodies capable of binding to the expressed proteins.
XX      The proteins can be used to screen compounds in rational drug discovery

```

```

CC      programmes. The antisense nucleic acid sequence is also useful to screen
CC      for homologous nucleic acids which are required for cell proliferation in
CC      a wide variety of organisms. The present sequence represents an
CC      essential prokaryotic cellular proliferation protein.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 477 AA;
XX
Query Match 4.4%; Score 116.5; DB 22; Length 477;
Best Local Similarity 20.2%; Pred. No. 0.0013;
Matches 95; Conservative 84; Mismatches 186; Indels 105; Gaps 27;

Qy      69 WLVPDCAITFMILYAKLNT-LSKQALFYAVGTPPLIFALPPTYIYPLRDVLHTEFPAD 127
Db      51 WISVPCLLAFVWMLFSAVVVNLPKVGFNFTDOLFMLTALPSVSGALLXVPY----- 104
Qy      128 RLQALTPPGLIGVAILNNMTPAARFYVLAELGWSVMSLMFWMFANEITKIHAKRFYAL 187
Db      105 --SKWVP--LFG---GRRWT--AFST-----GIIVPCWLGXA-----VDISTPST 144
Qy      188 FGIGANISLISGRAIVWASKLRASVS-----EGYDPMGISRLLMAMTI 232
Db      145 FII---ISLLCXXGAXFASSM-ANISFFPKQKGALGUNGGLGNNGVSVMLXAPLV 200
Qy      233 V-----SGIVMASYW-WIN-KNVLTDPRFVNPDEMOKGKAKPRM 272
Db      201 VLSIFAAGSHGVEQPDGSQLYLANAAMIWPFIAIFTLAMFGMNEI-----ATSKA 254
Qy      273 NKKDSFVLDRSPVYLLTLTLVIA-YGICINLEVTWKSQKLOYPNNDYSEPMGFNSF 331
Db      255 SLKEQLPVLRKG-HMWTMSLYLATFGSFIG-FSAGFAMLSKTQFPDVQ-----ILHTAF 307
Qy      332 WTGVVSVLIMLFEVGNVIRKGMLTGALVTPV-NVLLTGIVFPFALVIFRQASGLVAMEG 390
Db      308 RGFPIGALAR-SAGALISDRIGTRVTLINFLYMAIFSGLLFLTLPT-GVGGSFIATFG 365
Qy      391 TTPMLAVV-----GAIQNILSKSTYKALFDSTKEMAVYPLDQEKVKGAALDVVA 444
Db      366 --VFALFLTAGLSSGSTFQWISVIFRKLTMDRYKAGS---GSDERAMEBAADTTAA 418
Qy      445 -----RFGSGALIQOGL---LVICGS-IGAMPTYLAIVLLFTIAIMLV 485
Db      419 LGFISALGIGFPIPKAFGSSALRTGSPVGAMKVFLIFYACVITWAV 468

RESULT 36
AAE30467
ID      AAE30467 standard; Protein; 635 AA.
XX
AC      AAE30467;
XX
DT      24-FEB-2003 (first entry)
XX
DE      Haemophilus influenza cytochrome C-type biogenesis protein.
XX
KW      Virulence; veterinary; infection; pneumonia; antimicrobial drug; vaccine;
KW      antibiotic; gene therapy; cytochrome C-type biogenesis protein;
KW      antibacterial.
XX
XX      Haemophilus influenzae.
XX
OS      Haemophilus influenzae.
XX
PN      WO200277020-A2.
XX
PD      03-OCT-2002.
XX
PF      18-MAR-2002; 2002WO-GB01305.
XX
PR      22-MAR-2001; 2001GB-0007234.
XX      23-MAR-2001; 2001GB-0007360.
XX

```

PA (ISIS-) ISIS INNOVATION LTD.
 XX Herbert MA, Deadman ME, Hood DW, Moxon ER;
 PI WPI: 2003-029913/02.
 XX DR N-PSDB; MAD47835.
 XX DR N-PSDB; MAD47835.
 PT New virulence gene from Hemophilus influenzae, useful for producing
 PT vaccines or antibiotics for preventing or treating pneumonia -
 XX
 PS Claim 4; Page 70-72; 132pp; English.
 XX
 CC The present invention relates to Hemophilus influenzae virulence genes
 CC and proteins encoded by them. The microorganisms or the peptides of
 CC the invention are useful for manufacturing a medicament for treating
 CC (veterinary) or preventing a condition associated with H. influenzae
 CC infection, particularly pneumonia or for identifying an antimicrobial
 CC drug. Sequences of the invention are useful in the production of vaccines
 CC or antibiotics to prevent or treat H. influenzae infection. They are
 CC also used in gene therapy. The present sequence is H. influenzae
 CC cytochrome C-type biogenesis protein.
 XX
 SQ Sequence 635 AA;
 Query Match 4.4%; Score 116; DB 24; Length 635;
 Best Local Similarity 19.9%; Pred. No. 0.0022;
 Matches 103; Conservative 46; Mismatches 14; Indels 226; Gaps 21;
 QY 78 FMIIYAKLSNLSKQALFYAVGTPPLIFPFIYVPELRDVLHPTBPADRLQALPPGL 137
 DB 8 FLILLATAS-----AFPLAL-VPOFGLFKKPTLI-----NAWPLSYIFLATTLSIGL 56
 QY 138 LG-----LVAILRMTPFAFYVLAEIEMGSMVLSMFNGFA----- 172
 DB 57 LAYFAVADFTLEYVAHAHNSQLPTEFFKVAATGHEGSMFMDLSLMLAFAFNRK 116
 QY 173 NEITRIHAKRFYALFGIGANISL-----ASGR----- 201
 DB 117 NDRFSAQSLSLGLICGFPAVFIYNSNPGRIFFPAABGRDNLPMLODVGILFHPPL 176
 QY 202 -----AIWASKIRA-----SVSEGVDPWGISLRLLMAMTVSGVLMA SY----- 242
 DB 177 YGVGVFAVNFAMSLALYNQASARQIRSRGWVLVWMLPLTIGIVLG-AMWAVYELGW 235
 QY 243 --WINKNVLTDPFPYNEEMQKKGAKKPKMKNKDSFLYIDRSPIYLLTLVIANGIC 300
 DB 236 GGMW-----FMDPVE----- 245
 QY 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIMLF-VGQNVIRKFGMLTGAL 359
 DB 246 -NASLMFPLGLALHLSLM--ATEKQGVFSYWTTLFSLAFAFSVLGTFIYRSGALISVH 302
 QY 360 V-----TPVWVLTGIVEFALVIFRNOASGLVAMEGTTPLMLAVVGAIONLSKSTKYA 414
 DB 303 AFALDNRGVYLL-LIEFVLTAL--AFGLFALRA----- 333
 QY 415 LFDSTKEMAYPLDQKQVAKAIDVYARF-GKSGGALLQGLLYVC----- 462
 DB 334 -----GSSSESARKFOFISKSGILLNLTITATVSTFLGTFFP 373
 QY 463 -----GSIGAMPYLL-----AVILLFTIAT 482
 DB 374 MLFQAMWMSISVGSPIYNSIFPPIITAILIMVIVL 410
 RESULT 37
 AAG90107
 ID AAG90107 standard; Protein; 410 AA.
 XX AAG90107;
 AC AAG90107;
 XX 26-SEP-2001 (first entry)
 XX

DE C glutamicum protein fragment SEQ ID NO: 3861.
 XX
 XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX EPI108790-A2.
 PN
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 XX
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159152.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAH65326.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Claim 17; SEQ ID NO: 3861; 246pp + Sequence listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 410 AA;
 Query Match 4.4%; Score 115.5; DB 22; Length 410;
 Best Local Similarity 22.8%; Pred. No. 0.0013;
 Matches 72; Conservative 47; Mismatches 104; Indels 93; Gaps 16;
 QY 157 ELMGSMVLSL-MFEGANEITKI--HEAKRFYA-----LFGIGANISLHAGS-- 200
 DB 20 KVMIAVALSVFTVAMG-GNETPLLVFVRBEGFSNLFIDLLVFAIGVAVGLAAGPL 78
 QY 201 -----RAIYASKLRASVSEGVDPWGISLRLLMAM-----TVSGVLVMA SYWINKNV 249
 DB 79 SDRYGRRAVWLPAPLIIALLSGALIASGETAILIAIGRVLSIGSVGMWMTAGSGMIKX-- 136
 QY 250 LIDPRFYNPEEMQKKGAKK-PKNNKDSFL-----YIDRSPIYLLTLTLV 294
 DB 137 LSSSRF--BPQVTSAGARASVSLTGFPALGALGAWAQMPLPQQLAVVHITLTL 193
 QY 295 IAVGICINLEVTWKSQKLQYPMNDYSEFMGNFSFMTGV--SVLIMLFVGNVIRKF 352
 DB 194 ILFPPLITAPETRSALHKT-----GSFMSDVIVPSALDKRFL--FVAPI 238
 QY 353 G-WLTGALVTPWVYL-----LTGIVFPALVIFRNOASG----- 384
 DB 239 GPWVGAAFTAYAVLPQLRDMVASAPVAYGALIALVTLGSGFGIQFQPGQIMGTSKTRGP 298
 QY 385 LVAMFGTTPMLAVVV 400

DR N-PSDB; AAD11172.

XX New isolated nucleic acids, useful for producing enzymes required to
 PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
 PT reducing the amount of metal or carbon tetrachloride in a substrate,
 PT e.g. soil or water

XX Claim 32; Page 100-101; 172pp; English.

XX The present sequence is a Pseudomonas stutzeri open reading frame-N
 CC (ORF-N) protein. The Pseudomonas stutzeri genome includes ORFs that
 CC encode enzymes required for biosynthesis of pyridine-2,6-bis
 CC (thiocarboxylate) (PDBC). The ORFs encoding PDBC are especially useful
 CC in environmental remediation methods, e.g. phytoremediation,
 CC bioaccumulation, water purification, waste water purification, solution
 CC mining mobilization, immobilization, detoxification, redox state
 CC modifier or modification of metal ion reactivity. In particular, the
 CC ORFs are useful for degrading carbon tetrachloride and removing metals
 CC from substrates, e.g. soil or water.

XX Sequence 392 AA;

Query Match 4.4%; Score 115; DB 22; Length 392;
 Best Local Similarity 20.9%; Pred. No. 0.0014;
 Matches 88; Conservative 55; Mismatches 127; Indels 152; Gaps 20;

OY 94 LFVANGTPLIFLFPVITVPLRDVLAHTERAD---RLQ-ALLPGLGLVALIRWT 148
 DB 18 LLIAMGPMWMTFYAI-----GILGPHLVADIGISROQLGWTASTFGLAALISPA 68
 OY 149 FA-----AFVYLAELMGSVMLSMFMFANETTKHEAKRFALFGIGANISL 196
 DB 69 GALVORMGTRAGLICMFLVIGLSFSLMAVLP--GGGLVTAL-----LLCGTASLAN 119
 OY 197 LASGAIWASKL-BASVSEGVDPWGISRLMAMTVISGLVLM----- 239
 DB 120 PATNOALIASVPAARRAGVGLKQSGVQASALLAGVALPVLVMGMGALAAWPAVALV 179
 OY 240 ----ASYWINKQVLTDRFVYPEEMQKKGAKPRAMKDSFLYLDSPYLLTLTVI 295
 DB 180 MAALVTYVPAKSV-----SAPSLPLR-----VSGPNVMTLSILMAI 215
 OY 296 AYGICINILEVTWKSQKLQYNNNDYSEFMGNFSFMTGV--VSYLIMLVFGNVIRKF-- 352
 DB 216 Q--LQAGLA-----LSFMTFLGAVIAQIGVSSTI-----GAMVSCGA 253
 OY 353 -GMLTGAIVTVM-----VLLTGIVF---FALVIFR--NOASGLVAMFGTTPMLAVV 399
 DB 254 MGIISRVLTLTPADKLKQETILLGLVFLIAGIALAVMEANTQGMPLMLGVTMGLTV- 312
 OY 400 VCAIONTLSKSTKYALFSDTKEMAYIPLDQEQVKGAIDVVAARFGKSGALIQGILL 459
 DB 313 --AASNAIAMS-----MLTRDRGFGGAATASAGML 339
 OY 460 VI 461
 DB 340 SV 341

RESULT 40
 AAG91270
 ID AAG91270 standard; Protein; 708 AA.
 XX AAG91270;
 AC
 XX 26-SEP-2001 (first entry)
 DT
 XX C glutamicum protein fragment SEQ ID NO: 5024.
 DE
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX Corynebacterium glutamicum.

XX EPI108790-A2.
 PN 20-JUN-2001.
 XX 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 98JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX (KYOWA) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH66489.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 5024; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 708 AA;

Query Match 4.4%; Score 115; DB 22; Length 708;
 Best Local Similarity 18.6%; Pred. No. 0.0034;
 Matches 106; Conservative 70; Mismatches 149; Indels 244; Gaps 25;

OY 109 PPTVIYPLRDVLAHTERADRLQAILPPG-----LGLVALIRNMTFAAFVLAEL 158
 DB 120 YPHVVRPARPAKPMPTSEKIMRGVAIGGVITVAGVLLVSAIQGM---LGPLGRV 175
 OY 159 WGSVWLSIMFMFANETTKHEAKRFYALF-----GIGA 192
 DB 176 IGAVLLAVLLIGAHHYAK--RGTRVEALVALTTSQIAPLATTSATIFILEMPPGIGS 233
 OY 193 ----NISLASGRAIYMA-SKLASVSEGVDPWGISRLMAMTVISG-----LVMA 240
 DB 234 LVAILIGNGLIVK--LMSLSKTEKSAEG-----HVFVAIAVSGSAILPALSA 284
 OY 241 SYWINKNVLTDRFVYPEEMQKKGAKPRAMKDSFLYLDSPYLLTLTVIAYGIC 300
 DB 285 DAWM-----PIRSIVAALLSYAIS 304
 OY 301 INLI-----EVTWKSQ-----LKLQYNNNDY 322
 DB 305 TNIRASWAPAVAILQFVLSAQMTWMPATIVGHTVALLVALTMDPFXITATDSHDI 364
 OY 323 S--EFMGNF-----SFWTGVSVULIMLVGAGNVIRKGMGLTGAIVTPMVLITGIVPAL 375
 DB 365 ALEBYMSFETNPVSTWGAVSPLYIVITTSMEIADVMPVLAIPACAVAAIGI--PAL 422
 OY 376 ----VIFRQASGLVAMFG-----TTPMLAV--VVGA----- 402
 DB 423 RSSDTASIEHQMRLLAVAGLAIATFVQLTFYGDLPFNLLVAVFLIAGALFPMILRM 482

```
Qy 403 -----IONILSKSTKY-----ALPDSTKEMAYIPLDQEQ 431
Db 483 LPPORQLGVPPVWAMLIAAVAMTGVLRLRVVISPLMLTDQALIQALLIVFI----- 536
Qy 432 KVKGAIDVVAARFGKSGGALIOGILVIT---CGSIGAMTPYL-----AVILLFII 480
Db 537 ---AATIQVRRSFYGHKLMQILVGLTILTSAISIVTITTFIGRLIAGNAGMMLGFII 592
Qy 481 A-----IWLVSATKL---NKLFLAQSAL 500
Db 593 GHATVSIILMWYIAALMLNRKLLDAFGAL 621
```

Search completed: November 25, 2003, 10:14:31
Job time : 59 secs

GenCore Version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: November 25, 2003, 10:16:48 ; Search time 404 Seconds
(without alignments)
3441.115 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKREKPFGLKRSFLWPIH.....AQSAKKEQVAVQEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPRO.epool/US09869433/runat_24112003.171458.18110/app_query.fasta_1.711
-DB=N Geneseq.19jun03 -QEMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=prco -NORR=ext -HEAPSIZE=500 MINLEN=0 -MAXLEN=2000000000
-USER=US09869433 @CGN_1_1_0 @runat_24112003.171458.18110 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq.19jun03.*
1: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2003.DAT.*
25: /SIDSI/gcgdata/geneseg/geneseg-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Match Length	DB ID	Description
1	2630	100.0	1637	AAA30922	C. pneumoniae ATP/
2	2617	99.5	273254	AA081914	Chlamydia pneumoniae
3	2617	99.5	133025	AA091990	Nucleotide sequenc
4	2141	81.4	1038602	AA0201425	Complete genome se
5	1333	50.7	1875	AB212889	Arabidopsis thalia
6	1330.5	50.6	2146	AA036046	Aradidopsis thalia
7	1306	49.7	1896	AA061374	Potato chloroplast
8	1260	47.9	1770	AA061373	A. thaliana chloro
9	1025.5	39.0	1823	AA084466	Chlamydia pneumoni
10	1025.5	39.0	1230025	AA091990	Nucleotide sequenc
11	256	9.7	284	AB174862	Corn tassal-derive
12	242	9.2	246	AB176356	Corn tassal-derive
13	242	9.2	280	ABX86909	Corn tassal-derive
14	227.5	8.7	309	AB171938	Corn tassal-derive
15	168	6.4	265	ABX84884	Corn ear-derived p
16	162	6.2	1614	AA065082	C glutaminc codin
17	162	6.2	349980	AA064966	C glutaminc codin
18	152.5	5.8	1038602	AA0201425	Complete genome se
19	144	5.5	1983	AA053363	Neisseria meningit
20	144	5.5	1983	AA053364	Neisseria meningit
21	144	5.5	2028	ABK54090	DNA encoding Neiss
22	144	5.5	60873	AA081469	N. meningitidis pa
23	144	5.5	349980	AA061610	Neisseria meningit
24	144	5.5	1437668	AA081490	N. meningitidis B
25	138.5	5.3	1389	AA052427	E. coli DNA for ce
26	137	5.2	1613	AB070453	Listeria monocytog
27	134.5	5.1	349980	AA061848	Brifidobacterium lo
28	134	5.1	1983	AA053362	Neisseria gonorrhoe
29	134	5.1	2025	AB239606	N. gonorrhoeae nuc
30	134	5.1	349980	AA061610	Pyrococcus abyssi
31	133	5.1	1830121	AA0742063	Haemophilus influe
32	132.5	5.0	1392	AA052410	E. coli DNA for ce
33	129.5	4.9	284458	AA0803041	Listeria monocytog
34	126	4.8	664707	AB067196	Listeria innocua c
35	126	4.8	3011208	AB069245	Listeria innocua D
36	125.5	4.8	1305	AB075037	Pathogen specific
37	123.5	4.7	7477	AA088789	B. subtilis bac DN
38	123	4.7	236589	AB090521	Genomic sequence o
39	122.5	4.7	68750	AA025887	Sorangium cellulos
40	122	4.6	2547	AB117965	Drosophila melanog
41	122	4.6	2564	AB118561	Drosophila melanog
42	121	4.6	2541	AB070945	Listeria monocytog
43	121	4.6	5482	ABK12653	Mouse voltage gate
44	120.5	4.6	1188	AA056157	Salmonella typhi D
45	120.5	4.6	1278	AA054436	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAA30922
ID AAA30922 standard; DNA; 1637 BP.

AC AAA30922;
22-SEP-2000 (first entry)

C. pneumoniae ATP/ADP translocase coding sequence.

ATP/ADP translocase: Chlamydia infection; diagnosis; therapy; ds.

OS Chlamydia pneumoniae.

XX Key Location/Qualifiers

FT CDS 51..1598
/*tag= a

/product= ATP/ADP_translocase

FT XX WO200039157-A1.
 XX PD 06-JUL-2000.
 XX PF 22-DEC-1999; 99WO-CA01224.
 XX PR 28-DEC-1998; 98US-0114060.
 XX PR 12-MAR-1999; 99US-0123967.
 XX PR 30-JUN-1999; 99US-0141271.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX DR WPI, 2000-452368/39.
 XX DR P-PSDB; AAV90265.
 XX PT Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
 XX PT prevention and treatment of Chlamydia infection in mammals -
 XX PS Claim 2, Fig 1; 81pp; English.
 XX CC This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase
 XX CC of the invention. The protein, DNA encoding it, or a vaccine containing
 XX CC the DNA or protein, are useful for diagnosing, preventing or treating
 XX CC Chlamydia infection. The sequences can also be used in a method for
 XX CC the detection of Chlamydia infection. Primers or probes derived from the
 XX CC DNA sequence are useful in diagnostic tests for detecting Chlamydia
 XX CC infection.
 XX SQ Sequence 1637 BP, 431 A, 327 C, 325 G, 554 T; 0 other;

Alignment Scores:
 Pred. No.: 1.81e-276 Length: 1637
 Score: 2630.00 Matches: 515
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-869-433-2 (1-515) x AAA03922 (1-1637)

QY 1 MetThrLysThrGluGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
 Db 51 ATGCAAAAGCGAAGAAACCTTTTGGAAATGCGCTCTTTCTTGCGCGATACAT 110
 QY 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 Db 111 ACTCAGAGCTTAAGAAAGTCTGCCAATGTTCTTAATGTTCTTCTGATTAACATTTTAC 170
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyVala 60
 Db 171 TATACGGGTGTAACGGATCAAAAGACACTTATGTTGGAGGCGCTCGGTTCGTGTCA 230
 QY 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 Db 231 GAGGCAATACCTTCATCAAGTTTGGCTTGTGCTTGGCTTGGCTATATATCTTATGCTT 290
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 Db 291 ATTATGCAAGAGCTAAGTAATTTTAAAGTAAGGAGGCTTATTTATGCAAGTGAACG 350
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 Db 351 CCGCTTTTAAATTTCTTCCCTGTTCCGACTGTAATTTATCCGCTACGGAGTGTTTTA 410
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 Db 411 CATCTCATGAGATTTGTGACCGCTTACAGGCGCACTTACCTCCAGATTCCTAGGACTC 470
 QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyrValLeuAlaGluLeuTrpGly 160

Db 471 GTTGCATCTTAAGAACTGACATTTGCTGCACTTTATGACTTGCTGAGCTATGGGA 530
 QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleTrpLysIleHisGlu 180
 Db 531 AGCGTCATGCTATCTTAATGTTCTGGGGATTTCTTAATGAATAATCAAAATCCACGAA 590
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 Db 591 GCAAGGCTTCTACGCTCTTTCGTATCGGACTAATATTTCTTACAGCTTCTGCT 650
 QY 201 ArgAlaIleValITrPalSerLysLeuArgAlaSerValSerGluGlyValAspProTrp 220
 Db 651 CGTGCATTTGTTGGCTTCGTAAGTGAAGTCCGTTCTGAGGATGATCCCTGG 710
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 Db 711 GGAATTTCTTACCGTCTTTCATGCTATGACTATGATCTGAGACTTGTCTTATGCGCC 770
 QY 241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu 260
 Db 771 AGTACTGATGATCAATGAAGACGATGACCGATCTGCTCTTAATCCAGAAAGA 830
 QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetLysAsnMetLysAspSerPheLeuTrp 280
 Db 831 ATGCAAAAGGGGAAAAAGGCTTAACCTTAATGAATGAAGATAGACTTCTCTAT 890
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuValIleAlaTyrGlyIleCys 300
 Db 891 CTGATATGATCTCCTTATATCTTTATTAATCTCTGTTATGATGCTATGATTTGC 950
 QY 301 IleAsnLeuIleGluValIThrTrpLysSerGlnLeuLysGlnTyrProAsnMetAsn 320
 Db 951 ATTACTTAATCGAAGTACTGGAAGATGAGTGAACCTGAACCTTAATGAATGAT 1010
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValAlaSerValLeuIle 340
 Db 1011 GACTATGATGATTCATGAGGGAATCTTCTCTCGAGCTGGCGTATGCTGATATC 1070
 QY 341 MetLeuPheValGlyIleAsnValIleArgLysPheGlyTyrPheLeuThrGlyAlaLeuVal 360
 Db 1071 ATGCTATTTGTTGGTGTGAACGTCATTCGTAATTTGATGATGCTGAGGCGCTTATGTC 1130
 QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
 Db 1131 ACTCTGTCAATGTTCTCTCAACGATATGTTCTGCTCTGTTATCTTTAGAAC 1190
 QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValAlaVal 400
 Db 1191 CAAGCTTCTGGGCTGCTGCTATGCTATGCTCAATCTCTCATGCTAGCTGTGTTGTC 1250
 QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
 Db 1251 GAGAGCTTACAGAAATATCTTTCGAAATCCAAATCAACCTCTCTTGTGCTCAACCTAAA 1310
 QY 421 GluMetAlaTyrIleProLeuAspGlnGlnLysValIleGlyLysAlaAlaIleAsp 440
 Db 1311 GAAATGGCTATATCCCTCTTTCGACCAAGACCAAAATCAAGGTAAAGGCTGCTATGAT 1370
 QY 441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuVal 460
 Db 1371 GTAGTCCCGCCCGCTTCGGAATAATCGAGAGCTTATCCACAGAGTTTCTGCTGTT 1430
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrIleuAlaValIleLeuLeuPheIle 480
 Db 1431 ATCTGTGAGATATTTGAGACTATGACCCCTTATCTGCACTGATCTTCTTTCATCAT 1490
 QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 Db 1491 GCTATTTGTTGTTCTTCTGCACTTAAGTTAAACCAATATTTCTTAAGCGCAGTGTGCTT 1550
 QY 501 LysGluGlnGluValAlaGlnGlnLysAspSerAlaProAlaSerSer 515

DB 1551 AAGAACAAGAGTGGCTCAAGAATGATTACGCTCCGTCTTTCA 1595

RESULT 2
ID AAC81914/c standard; DNA; 273254 BP.
XX
AC AAC81914;
XX
DT 27-FEB-2001 (first entry)
XX
DB Chlamydia pneumoniae genome DNA.
XX
KW Genome; diagnosis; vaccine; ds.
XX
OS Chlamydia pneumoniae.
PN WO200027994-A2.
PP 18-MAY-2000.
PD
PF 12-NOV-1999; 99WO-US26923.
PR 12-NOV-1999; 98US-0108279.
PR 08-APR-1998; 98US-0128606.
XX
XX (REGC) UNIV CALIFORNIA.
PA
PI Stephens R, Mitchell W, Kaiman S, Davis R,
DR MPI, 2000-376516/32.
XX
PT Isolated nucleic acid for use in diagnostic and analytical methods
PT encodes genomic sequence of Chlamydia pneumoniae -
PS
XX Claim 2; Page 128-320; 320pp; English.

This invention describes a novel nucleic acid (N1) encoding a Chlamydia pneumoniae protein (P1), given in the specification. The isolated nucleic acid is useful for diagnostic and analytical methods, such as, hybridization-based assays or amplification-based assays. The protein may be used for diagnostic purposes, for their enzymatic or structural activity, or as a vaccine. The invention also describes (1) a probe comprising a hybridizing fragment of N1; (2) an isolated nucleic acid (N2) that hybridizes under stringent conditions to N1; (3) an expression cassette comprising N1 under the transcriptional regulation of a transcriptional initiation region functional in an expression host, and a transcriptional termination region; (4) a cell comprising an expression cassette of (3) as part of an extrachromosomal element or integrated into the genome of a host cell as a result of induction of the expression cassette into the host cell, and the cellular progeny of the host cell; (5) a method for producing a P1 comprising growing a cell of (4) where the protein is expressed and isolating the protein free of other proteins; (6) a purified polypeptide composition comprising at least 50 weight % of P1; and (7) a monoclonal antibody binding specifically to the peptide of (6).

SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Alignment Scores:
Pred. No.: 8,43e-272 Length: 273254
Score: 2617.00 Matches: 513
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 2
Query Match: 99.51% Indels: 0
DB: 21 Gaps: 0

US-09-869-433-2 (1-515) x AAC81914 (1-273254)

OY 1 MetTrIrrysThGlGluLysFroPhcGIYLysLeuAArgSerPheLeuTPProIleHis 20
Db 212420 ATGCACAAAACCGAAGAAATAACTTTTGAAAAATTTCGCTCTTTCTTGCGCGGAATCAT 212361
OY 21 ThrHISgluleuLyslyValleuProMetPheLeuMetPhePheCysilethrPheasn 40

Db	212360	ACTCAGCGACTTAAGAAAAGTTCTGCCAATGTTCTCTTAATGTTCTTCTGTAATTACATTTAAC	212301
QY	41	TYRTHValLeuArGAspThrLySAspThrLeuIleValAlaProGlyLeuSerGlyAla	60
Db	212300	TATACGGGTGTTACGGATACAAAAGACACTTTATGTGGAGCTCCTGGTTCGGTGCA	212241
QY	61	GIuAlaIleProPheIleLeuPheTrpIleuValProCysAlaIleIlePheMetLeu	80
Db	212240	GAGGGAATACCTTTCATCAGATTTGGCTGTGTGCCCCCTGGCTAATATCTTATGCTT	212181
QY	81	IleTryAlaLySLeuSerAsnIleLeuSerLySGlnAlaLeuPheTryAlaValAlaGlyThr	100
Db	212180	ATTATGCAAGCTTAGTAATATTTTAAGTAGCAGGCGCTTATTTTATTCAGAGTGGAAAG	212121
QY	101	ProPheLeuIlePhePheAlaLeuPheProThrValIleTryProLeuArGAspValLeu	120
Db	212120	CCCTTTTAATTTTCTTTGGCCCTGTTCGCACTGTATATTTATCCGCTACGGATGTTTTTA	212061
QY	121	HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu	140
Db	212060	CATCCTAAGAAATTTGCTGACCGGTGTACAGGCACTCACTCCAGGATTTGCTAGACATC	212001
QY	141	ValAlaIleLeuArGAsnTrpThrPheAlaAlaPheTryValLeuAlaGluLeuTrpGly	160
Db	212000	GTTGCACTTTAAGAACTGACATTTGCTGATTTTATGTAAGTCTGCTAAGCTATGAGGA	211941
QY	161	SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLySLeuIleGln	180
Db	211940	AGCGTCAGCTATCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAANAATCCAGAA	211881
QY	181	AlaLySArGheTryAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly	200
Db	211880	GCAAAAGCTTTCTACGCTCTTTCCGTAACGAGCTAATATTTCTTACTAGCTTCGTGT	211821
QY	201	ArgAlaIleValTrpAlaSerLySLeuArGAlaSerValSerGluGlyValAspProTrp	220
Db	211820	CGTGAAATGTTGGGCTTCAAAAGTGAAGGCTTCGTTCTGAAGGTGAAATCCTGTGG	211761
QY	221	GlyIleSerLeuArGLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240
Db	211760	GGAAATTCCTTAACGCTCTTTGATGGCTATGACTATTTGTAATCTGGACTTTTCTTAAGCC	211701
QY	241	SerTryTrpTrpIleAsnLySAsnValLeuThrAspProArGheTryAsnProGluGln	260
Db	211700	AGTACTGCTGGATCAATAGAAGCTATGACCGATCCCTCGCTCTATATCCAGAAAGA	211641
QY	261	MetGlnLySGlyLySLeuGlyValAlaLySProLySMetAsnMetLyAspSerPheLeuTry	280
Db	211640	ATGCAAAAAGGGGAAAAAAGGGCTTAACCTAAAGAATATAAAGATGCTTCCTCTAT	211581
QY	281	LeuAspArGSerProTryIleLeuLeuLeuThrLeuLeuValIleAlaTryGlyIleCys	300
Db	211580	CTTGCTAGATCTCCTTAATATTCCTTTATTAAGCTCTCTTGAGTTATTCCTTAATGTAATTTGC	211521
QY	301	IleAsnLeuIleGluValThrTryLySArgLeuLeuLySLeuGlnTryProAsnMetAsn	320
Db	211520	ATTAACTTAATCGAAGGACTTGGAAGAGTCAAGTCAAACTCAATATCTTAATATGAAT	211461
QY	321	AspTrySerGluPheMetGlyAsnPheserPheTrpThrGlyValAlaSerValLeuIle	340
Db	211460	GACTATATGATGATTCATGAGGGAACTTCCCTTCGACCTGGGCGTGTATCCGACTATTATC	211401
QY	341	MetLeuPheValAlaGlyAsnValIleArgLySPhesGlyTrpLeuThrGlyAlaLeuVal	360
Db	211400	ATGCTATTATGTTGGTGTACCTCAATTCGTAAATTTGGATGGATTAATCGAGCCCTAGATC	211341
QY	361	ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn	380
Db	211340	ACTCGTGTCAAGGTTCTCTCTAACAAGATATCGTTTTCTTGCGCTTGTTATCTTTAGAAAC	211281
QY	381	GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal	400

Db 211280 CAAGCTTCTGGGCTGTCGATGTTGGTACCACTCCTCATGCTAGCTGTGTC 211221
 QY 401 GYAlaAlleglnaenlleuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
 Db 211220 GGAGCTATACGAATATCTCTTCGAAATCCAAATACGCTCTTGGACTCACTAA 211161
 QY 421 GlnuSerLysTyrIlePheLeuAspGlnGlnGlnValLysGlyLysAlaAlaIleAsp 440
 Db 211160 GAATGGCTATATCCCTCTTGACCAAGCAAAAGTCAAGGTAAGGCTGAT 211101
 QY 441 ValValAlaAlaArgPheGlyLysSerGlyAlaLeuIleGlnGlnLysLeuVal 460
 Db 211100 GTAATGCTGCTCCGCTTCGAAATACGAGAGCTTTAATCCAAAGTTGCTGCT 211041
 QY 461 IleGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuPheIle 480
 Db 211040 ATCTGTGAAGTATGGAGCTATGACCCCTTATCTTGCAAGTATCTTTCATCAT 210981
 QY 481 AlaIleThrLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 Db 210980 GCTATTTGGTTGGTTCTTGCACTTAAGTTAAACAAATATCTTAAGCGCACTGCTCT 210921
 QY 501 LysGlnGlnGlnValAlaGlnGlnLysAspSerAlaProAlaSerSer 515
 Db 210920 AAGAACAAGAGTGGCTCAAGAAAGATTCACTCTGCTTCTCA 210876
 RESULT 3
 AAX91990
 ID AAX91990 standard; DNA; 1230025 BP.
 XX AAX91990;
 AC
 XX 13-SEP-1999 (first entry)
 DT
 XX
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 PA (GENSET) GENSET.
 PI Griffiths R;
 XX
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis, and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Alignment Scores:
 Pred. No.: 7.61e-271 Length: 1230025
 Score: 2617.00 Matches: 513
 Percent Similarity: 99.61% Conservative: 0
 Best Local Similarity: 99.61% Mismatches: 2
 Query Match: 99.51% Indels: 0
 DB: 20 Gaps: 0
 US-09-869-433-2 (1-515) x AAX91990 (1-1230025)
 QY 1 MetThrLysThrGlnGlnLysProPheGlyLysLeuArgSerPheLeuTyrProIleHis 20
 Db 404370 ATGCAAAAACCAAGAAAACCTTTGAAAATGGCGCTCTTCTGTGGCGCATCAT 404429
 QY 21 ThrHisGlnLeuLysLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 Db 404430 ACTCACGAGCTAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGTATTAATTAAC 404489
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
 Db 404490 TATACGGTGTACGCGATACAAAAGACACTTATATGTGGAGCTCTGTGTTCTGGTGA 404549
 QY 61 GlnAlaIleProPheIleLysPheTyrLeuValValProCysAlaIleIlePheMetLeu 80
 Db 404550 GAGGCAATACCTTTCATCAAGTTTGGCTTGTTGTCCTCCCTGTGATATCTTAATGCTT 404609
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 Db 404610 ATTATGCAAGCTTAAGTATATTTTAATGATACAGGCTTATTTATGACGTGGGAACG 404669
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 Db 404670 CCTTTTAATTTCTTGGCCGTGCTCCAGCTAATTTATCCGCTACCGAGTTTGA 404729
 QY 121 HisProThrGlnPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuGlyLeu 140
 Db 404730 CATCTTCACAAATTTGCTGACCGTTTACAGGCCATCTTACTCCAGATTGCTAGACATC 404789
 QY 141 ValAlaIleLeuArgAsnTyrPheAlaAlaPheTyrValLeuAlaGlnLeuTyrGly 160
 Db 404790 GTTGCCATCTTAAGAAACGAGACATTGCTGCAATTTATGATGCTGTAACATAGGGA 404849
 QY 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGlnIleThrLysIleHisGln 180
 Db 404850 AGGCTATGCTATCTCAAGTCTGGGAGTTGCTATATATAATCAAAAATCCAGAA 404909
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuAlaSerGly 200
 Db 404910 GCAAAGGTTTCTACGCTCTTTCCGATCGAGCTATATTTCTTACTAGCTTCGT 404969
 QY 201 ArgAlaIleValTyrPalaserylLeuArgAlaSerValSerGlnGlyValAspProTyr 220
 Db 404970 CGTGCAATTTGTTGGGCTTCAAGTTGAGAGCTTCCCTTCTGAAGGTGATGCCCTTGG 405029
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 Db 405030 GGAATTTCTTACGCTTTTGAGTGCGTATGACTATGTATCTGACTGTTCTTATAGCC 405089
 QY 241 SerTyrTrpTyrIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGlnGlu 260
 Db 405090 AGTTTACGTGGATCAATTAAGAACGTATTAACGATCTCTGCTTCTATATCCAGAA 405149
 QY 261 MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyr 280
 Db 405150 ATCAAAAAGGGGAAAAAGGTGCTAAACCTAAATGAATGAAGAAGTACTCTCTAT 405209
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 Db 405210 CTGTGTAAGATCTCTTATATCTTTATATGCTCTCTGTATATGCTATGGAATTTGC 405269
 QY 301 IleAsnLeuIleGluValThrTyrLysSerGlnLeuLysGlnLysProAsnMetAsn 320

```

Db 405270 ATTAACCTTAATCGAAGTACTGGAAAAAGTCAGCTGAACCTGAAATATCTTAATATGAAAT 405329
QY 321 AsPTyTserGluPheMetGlyAsnPheserPheThrPhgGlyValValSerValLeuIle 340
Db 405330 GACTATATGATGATGAGGAGAACTTCTCTTGAAGTCGCGGATGATCCGATCTATATC 405389
QY 341 MetLeuPheValGlyGlyAsnValIleArgPhePheGlyThrPheLeuThrGlyValValLeuVal 360
Db 405390 ATGCTATTTGTTGGTGTGATACGATTCGTAATTTGGATGCTTAAGTACGAGCCCTTATGTC 405449
QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheValLeuValIlePheArgAsn 380
Db 405450 ACTGCTCATGTTGCTCTTCAACAGGATGCTTTCTTGCTGCTTGTATCTTAAAGAAC 405509
QY 381 GluAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal 400
Db 405510 CAAGCTTCTGGGCTGCTGCTGATGTTGGTAACTCTCTCATGCTAGCTGCTGCTGCTGTC 405569
QY 401 GValAlaIleGluAsnIleLeuSerLeuSerThrIleThrValAlaLeuPheAspSerThrIle 420
Db 405570 GGAGCTATACGAAATATCTTTTCAAAATCCAAATAAGCTCTCTTGACTCACTAA 405629
QY 421 GluMetAlaThrIleProLeuAspGlnGlnIleValIleValIleValIleValIleValIle 440
Db 405630 GAATATGCTATATCCCTCTTGAACCAAGCAAAAGTCAAGTAAAGGCTGCTATTTGAT 405689
QY 441 ValValAlaAlaArgPheGlyValSerGlyGlyValAlaLeuIleGlnGlnIleValLeuVal 460
Db 405690 GTAGTTCGCCCTCCCTCGGAAATCAGAGAGACCTTAATCCAAAGGTTGCTGCTGCT 405749
QY 461 IleCysGlySerIleGlyAlaMetThrProThrIleValIleValIleLeuPheIleIle 480
Db 405750 ATCTGTGAAAGTATTTGAGCTATGACCCCTTATCTTGAAGTATTTCTTTTCAAT 405809
QY 481 AlaIleThrLeuValSerAlaThrIleValLeuValIleValLeuValIleValIleValIle 500
Db 405810 GCTATTTGTTGTTGTTCTTGAACCTTAAGTAAACAACTATTTCTTGAAGGCTGCTGCT 405869
QY 501 LysGlnGlnIleValAlaGlnGlnIleValIleValIleValIleValIleValIleValIle 515
Db 405870 AAAGAAACAAAGAGCTCAAGAAAGATTCAGCTCTCTCTCTCTCA 405914

```

RESULT 4

AAZ01425
ID AAZ01425 standard; DNA; 1038602 BP.

AC AAZ01425;

DT 07-OCT-1999 (first entry)

DE Complete genome sequence of Chlamydia trachomatis.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KM paratrachoma; inclusion conjunctivitis; genital disease; peritrichitis;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KM bartolinitis; pneumopathy; venereal lymphogranulomatosis; ss.

XX Chlamydia trachomatis.

XX OS

XX MO9928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98MO-IB01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97PR-0015041.

XX PR 17-DEC-1997; 97PR-0016034.

XX PA (BEST) GENSET.

XX PI Griffiths R;

```

DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal urethritis;
CC epididymitis, cervicitis, salpingitis, peritrichitis, bartolinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

```

US-09-869-433-2 (1-515) x AAZ01425 (1-1038602)

QY 1 MetThrIleThrGlnGluValIleProPheGlyValLeuValIleValIleValIleValIleValIle 20

Db 392407 ATGACTCAACCCCGGAAAAACCTTTGGAAATTTGGCTCTTCTTCCGCGATACAC 392466

QY 21 ThrIleGluLeuValIleValIleProMetPheLeuMetPhePheCysIleThrPheAsn 40

Db 392467 ATGCATGAGCTGAAGAAAGTTCGCCAATGTTCTTAATGTTCTTCTGATATTTCAAT 392526

QY 41 TyrThrValLeuArgAspThrIleValIleValIleValIleValIleValIleValIleValIle 60

Db 392527 TACACGATTTTGAAGATACAAAGATACCTTATGTTACAGACCGGATCTGGACA 392586

QY 61 GluAlaIleProPheIleValIleValIleValIleValIleValIleValIleValIleValIle 80

Db 392587 GAGGCCATTCCTTCAATTAAGTGGCTAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 392646

QY 81 IleThrAlaIleProPheIleValIleValIleValIleValIleValIleValIleValIleValIle 100

Db 392647 ATCTACGCCAAGCTTAACCAATATTTTGAACAAACAGGCTTTTCTTGGAGGCTCTCA 392706

QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleThrProLeuArgAspValIle 120

Db 392707 CCAATGCTGATATTTCTTGGACGCTTCCGCTGATATACCTTGGCGATATTTCTT 392766

QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140

Db 392767 CACCACAACGCTTTCGCTGATCTACCAATCAATCTTCTCTCCGATTTATGGGATTC 392826

QY 141 ValAlaIleLeuArgAspThrPhePheAlaPheThrValIleValIleValIleValIleValIle 160

Db 392827 ATTCGATGCTACGACATGACATTTGCTGTCTACAGCTTGTGAACTTGGGGA 392886

QY 161 SerValMetLeuSerLeuMetPheThrGlyPheAlaAsnGluIleThrIleIleGlu 180

Db 392887 AGCGTATATGCTCTTCTTGAAGTCTGGGCTTGGCCAAATTAATTAATAATTAGGGA 392946

QY 181 AlaIleArgPheThrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200

Db 392947 GCTAAACGTTTCACGCTCTTCTTGGAGAGGACTAACCTACCTCTTAAATTTCTGCT 393006

QY 201 ArgAlaIleValIleThrAlaSerIleValArgAlaSerValSerGluIleValIleProThr 220

Dd	393007	CCAGCTATTATTCGGTCTTTAAATTACGTCGTAACGCTTGCGCAAGCCGTGAACCATGG	393066
Oy	221	GjIleSerLeuArgLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240
Dd	393067	GGAGTCAGCCTCATTTTCTTAATGGCTAATGTTCTTGAGTTCTTGCCCTATTATGGCTGC	393126
Oy	241	SerYTrpTrpTrpIleAsnLysAsenValIleuThrAspProArgPheYTrAsnPcGluglu	260
Dd	393127	TGCTACTGTGTGATGAACCGCTACGCTTACAGATCCAGATTCTACATCTCTGACAG	393186
Oy	261	MetGlnLysGlyLysLysGlyAlaLysProLysMetAsmMetLysAspSerPheLeuYr	280
Dd	393187	CTC---AAGCTAAGAAA--TCTAAACCTAAGATGACGTGGCGAAGCTTCTCTAT	393240
Oy	281	LeuAspArgSerProYrIleLeuLeuLeuThrLeuLeuValIleAlaYrGlyIleCyS	300
Dd	393241	CTGTTAAAGATCTCTTATATAGTCTTCTTAGAGCTTCTAGTTATTGTCATACGGAATCTGG	393300
Oy	301	IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnYrProAsnMetAsn	320
Dd	393301	ATTAACTCTGTGAAGTAACCTGGAAAAACCACTCAAATGCAATTCCTTATCCAAAT	393360
Oy	321	AspYrSerGlnPheMetGlyAsnPheSerPheTrpThGlyValValSerValLeuIle	340
Dd	393361	GACTACAGCGCAATTATAGGGAACTCTCTTCTTGACAGGAATGTATCTGTAATTGTA	393420
Oy	341	MetLeuPheValGlyGlyAsnValIleArgLysPheGlyYrPheuThrglyAlaLeuAl	360
Dd	393421	ATGCTCTTCATCGGTGTGTAACGTATACCGTAGATTGGTTGGTTAATCGAGCTTTGGTT	393480
Oy	361	ThrProValMetValIleuLeuThrglyIleValPhePheAlaLeuValIlePheArgan	380
Dd	393481	ACACCTATCATGTGTTCTCGTAAACAGAGCGTGTCTTGTGCTCTCGTTATCTTCAACAGAT	393540
Oy	381	GlnAlaSerGlyLeuValAlaMetPheGlyYrThrProLeuMetLeuAlaValVal	400
Dd	393541	CATGCTACAGGTTTAGTGTCTGCTTTGGGAACAACACCACATAATAGCTAGCGTTGTTGC	393600
Oy	401	GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyraAlaLeuPheAspSerThrLys	420
Dd	393601	GGAGCGATTCAAAATATTTCTTCCAATCAACATAATATAGCTCTCTTGATGCACTAA	393660
Oy	421	GluMetAlaTyrlleProLeuAspGlnGluInLysValLysGlyLysAlaAlaleasp	440
Dd	393661	GAGCTGGCTTACATCCCATTTGATCAAGACAAAAGTTAAAGGAAGACCTGCTATTGAC	393720
Oy	441	ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnInGlnGlyLeuLeuVal	460
Dd	393721	GTTGTTCGCCGTAGATTGGTAAATCTGGGGGATCTTTATCCAACAAAGTCTCTTGGTC	393780
Oy	461	IleCyGlySerIleGlyAlaMetThrProYrLeuAlaValIleLeuLeuPheIleIe	480
Dd	393781	CTTTGGGAGAAGATCGGCCCTATGACTCCATTTCTAGCTGTGCTCTCTTGGCAATTATC	393840
Oy	481	AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu	500
Dd	393841	ATGTCCTGGTGAATCTCACTCACTAACTAACTAACTGTTCTTGCTGATCCGCTGCT	393900
Oy	501	LysGluGlnGluVal-----AlaGlnGlu	508
Dd	393901	AAAAGAACAGGAATTAGNNAAGCTGCAGACGACGAGAGAGAAAGCTTCTTCTGCGGCTTAA	393960
Oy	509	AspSerAlaProAla	513
Dd	393961	GAATCCGCTCTGCT	393975
RESULT 5			
ID	ABZ12889	standard; DNA; 1875 BP.	
XX	ABZ12889;		
DT	21-JAN-2003	(first entry)	

```

xx Arabidopsis thaliana stress regulated gene SEQ ID NO 694.
DE Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX MO200216655-A2.
PN WO200216655-A2.
XX 28-FEB-2002.
PD 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26685.
PF 24-AUG-2001; 2001WO-US26685.
PR 24-AUG-2000; 2000US-227866P.
XX 26-JAN-2001; 2001US-26467P.
PR 22-JUN-2001; 2001US-300111P.
XX (SCRI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Harper JF, Kreps J, Wang X, Zhu T,
PI WPI; 2002-104127/34.
DR WPI; 2002-104127/34.
XX Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT -
PS Claim 144; SEQ ID NO 694; 577bp + Sequence Listing; English.
XX The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stresses. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX SQ Sequence 1875 BP; 466 A; 416 C; 455 G; 538 T; 0 other;
SQ Alignment Scores:
Pred. No.: 5_93e-135 Length: 1875
Score: 1333.00 Matches: 259
Percent Similarity: 69.60% Conservative: 89
Best Local Similarity: 51.80% Mismatches: 140
Query Match: 50.68% Indels: 12
DB: 24 Gaps: 3
US-09-869-433-2 (1-515) x ABZ12689 (1-1875)
QY 16 LeutrPrroIleHtiThriSiGlueUlySvVallEuprOmeChelEuNetPhieNe 35
Db ATTTTGGTGAGGATTTCCTTAACCAATCTTGAATAAAGATTATCCTTAGATTGATGTCTTT 3729
QY 36 CyellEtInrhEAstTyThrValLeuArGaRpnTrlvsAsGrThreuleileValGLyala 55
Db TGATATCTTTTCAATTAACAATTCGTGAGGGATACAAAGAATGCTTGtggTgGACGCc 4329
QY 56 ProGiJSeSrGIYAIGLualalleProphellelYSPheTripleValValProCYaLA 75
Db AAAGAAAGTTCGCGAGATTAATACCTTTCTTGAAGACTGGGTGAATCTTCTTAAGGCC 4922
QY 76 llellePheMetLeuileTyrrAlalyseUsSerAanilleLeuserylsglnalaleupe 95
Db ATTGGATTATATCTCTCTAACCTTAACCTTCCAATGTTCTCTCCAAGAAGCTCTGTTT 5522

```


QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
 DB 553 TACACTGTTATGTCCTTTCATCATCTTGGGGCTTTGGTTTGTCATGATACCT 612
 QY 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuAlaIleLeuProPro 135
 DB 613 CTCGACCACTATTATCCACCCGAGACCTCGCAGATTAAGCTCTTACCAACCTCGGCCCA 672
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTpphrPheAlaAlaPheTyrValLeu 155
 DB 673 AGATTCACTGGGTCCTATGCAATATTTGGGATTTGAGATTTCTGTTGTTTATGTTATG 732
 QY 156 AlaGluLeuTyrPGLySerValMetLeuSerLeuMetPheTyrPGLyPheAlaAsnGluIle 175
 DB 733 GCTGAGCTTTGGGGAGTGTGTGTCTCAGTCTCTTCTGGGGCTTTGCTTATGATC 792
 QY 176 ThrIleHisGluAlaIleAspArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 DB 793 ACAACTGTGATGAGCCAGCAAGAAATTCATCTTGTTCGAGCTTGAGCCCATGTTCA 852
 QY 196 LeuLeuAlaSerGlyArgAlaIleValTrrpAlaSerIleuArgAlaSerValSerIleu 215
 DB 853 CTGATTCTCAGAGAAAGCCGTGAATACTTCTTAACCTGAGAAAGAAATCTGTGCT 912
 QY 216 GlyValAspProTrrpGlyIleSerLeuArgLeuMetAlaMetThrIleValSerGly 235
 DB 913 GAGCTTGACGGCTGGGCACTTCCTTGAAGCAAGATGACATTTGTGTGGATGGA 972
 QY 236 LeuValLeuMetAlaSerTyrTrrpIleAsnIleValIleuThrAspProArgPhe 255
 DB 973 CTCGCCATTGTCCTCTATGTTGGGTGATGATATGTTCTTCCCAACCGTATGC 1032
 QY 256 TyrAsnProGluGluMetGlnGlyIleGlyIleValIleProIleMetAsnMetIle 275
 DB 1033 AAGAACAGAGAGAG-----AAACCGAATAGGAGACGATG 1068
 QY 276 AspSerPheLeuTyrLeuAspArgSerProTrrpIleLeuLeuThrLeuValIle 295
 DB 1069 GAAAGCTTGAAGTCTTGTGTATCATCACCATCATTAAGATCTTGCTACTTGTGTG 1128
 QY 296 AlaTrrpGlyIleCysIleAsnLeuIleGluValIleThrTrrpIleSerGlnLeuIle 315
 DB 1129 GCATACGGTATTAGTATCATCTTGTGAGACATGAGATCAAAAGCTTAAAGCTCAG 1188
 QY 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrrpIleVal 335
 DB 1189 TTCCTTGGCCGAAATGAGTACTCAGCATTTATGGAGACTTCTCAACCTGACGGGTCTT 1248
 QY 336 ValSerValLeuIleMetLeuPheValGlyIleAsnValIleArgIlePheGlyTrrpLeu 355
 DB 1249 GCACATTCACATGATGCTT---CTCAGCCAATACGATTCATAGTATGTTGGGGA 1305
 QY 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
 DB 1306 GTAGCTGCAAGATCACCCCAAGCTGTTCTGCTATGATGATGTTGGCTTCTTCTCTA 1365
 QY 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
 DB 1366 ATATGTTTGGCGGCCCATTCGACCACTTGTTCGCAAGCTTGATGACACCGTACTT 1425
 QY 396 LeuAlaValValAlaGlyAlaIleGlnAsnIleLeuSerIleSerThrIleTyrAlaLeu 415
 DB 1426 GCAGCTGTGATGTCGGTCCCTTCAGATTCCTTCAGCAAGATGCCAAGTACAGCTTG 1485
 QY 416 PheAspSerThrIleGluMetAlaTrrpIleProLeuAspGlnGluIleValIleGly 435
 DB 1486 TTCGACCTTCGCAAGAAATGGCTATATCCATTCATGATGAGACACCAAGATTAAAGGC 1545
 QY 436 LysAlaAlaIleAspValValAlaAlaArgPheGlyIleSerGlyIleValAlaLeuIleGln 455
 DB 1546 AAACCTGGATTTGACGTGTGTGCAACCATTAAGGAAATAGGGGAGCTTTATATACG 1605
 QY 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTrrpIleAlaValIle 475

DB 1606 CAGTTCATGATCTTATCCTTGGATCATCTACGAAATTCAGAGCCGTATCTGAAATATC 1665
 QY 476 LeuLeuPheIleIleAlaIleTrrpLeuValSerAlaThrIleAsnIleLeuPheLeu 495
 DB 1666 TTGTTGTTATTTGTCATGCGTGTGTAGCTCAGCTAAGTGTGTGAGAGACAGTTTC--- 1722
 QY 496 AlaGlnSerAlaLeuIleGluGlnValAlaGlnIleuAspSerAlaProAlaSerSer 515
 DB 1723 -----AACACTTGGCTCTGAAAGAGCTTGAAGAGAAATGAGAGAGCTTCATCG 1776
 RESULT 6
 AAC36046
 ID AAC36046 standard; DNA; 2146 BP.
 AC AAC36046;
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12350.
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 EN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 PE
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126254.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0151065.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149929.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156459.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158237.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159229.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160777.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 1,35e-134
Score: 1330.50
Percent Similarity: 68.88%
Best Local Similarity: 50.88%

Length: 2146
Matches: 260
Conservative: 92
Mismatches: 136

Query Match: 50.59% Indels: 23
 DB: 21 Gaps: 5
 US-09-869-433-2 (1-515) x AAC36046 (1-2146)

QY 16 LeuTrpProIleHisThrHisGluLeuValysValLeuPrometPheLeuMetPhePhe 35
 DB 416 ATTTTCGGTGTGGAGGTTAGACACTGAGAGAGATGTCTTCTTAGGGCTTAATGTTCTT 475
 QY 36 CysIleThrPheAsnYrThrValLeuArgAspThrIlysAspThrLeuIleValGlyAla 55
 DB 476 TGCATCTTTTCATTTACACAACTCTTAGGACACGAAAGATGTTTGTGTGATGACGGCT 535
 QY 56 ProIysSerGlyAlaGluAlaIleProPheIleYsPheThrPheLeuValProCysAla 75
 DB 536 AAAGCAAGTTCTGCTGAGATTAATACCTTTTGAAGACATGGGAAATCTTCCATGGCT 595
 QY 76 IleIlePheMetLeuIleYrAlaIleYsLeuSerAsnIleLeuSerIysGlnAlaLeuPhe 95
 DB 596 ATTGGGTTTATGTTGCTAATACCAACTTCCAAATGTTCTCCAAAAGGCTCTCTT 655
 QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleYrPro 115
 DB 656 TACACTGTATGTTCTTCTTCAATGTTCACTTGGAGCCTTGGTTGCTGATGATCCCT 715
 QY 116 LeuThrArgPheValLeuHisProThrGluPheAlaIleArgGlnAlaIleLeuProPro 135
 DB 716 CTCAGCAATTTGATTCATCTGAAAGCTCTTCTTAAGCTTCTTGCAACACTGGCCCC 775
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnYrThrPheAlaIlePheYrValLeu 155
 DB 776 AGATCATGGCTCTCTCGCAATCATAGAGATTTGGAGTTCTGTTTCTTAATGTCATG 835
 QY 156 AlaGluLeuYrGlySerValMetLeuSerLeuMetPheYrGlyPheAlaAsnGluIle 175
 DB 836 GCTGAGCTTGGGGTAGTGTTGTTCACTTCTCTGGGATTTGGCAACACAGATT 895
 QY 176 ThrIleIleHisGluAlaIleYsArgPheYrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 DB 896 ACAACTGTGACGAAGCAAAAGTTCTATCTCTGTTGACCTGGGGCAAAAGTGCA 955
 QY 196 LeuLeuAlaSerGlyArgAlaIleValYrAlaSerIysLeuArgAlaSerValSerGlu 215
 DB 956 CTTATCTTCTCAGAGAAAGCTGTAATTTCTTAATAGAGAAAGAACTTGTCCT 1015
 QY 216 GlyValAspProYrGlyIleSerIleuArgLeuLeuMetAlaMetThrIleValSerGly 235
 DB 1016 GSAAGTTGATGGCTGGCTGTTTCACTTAAGCTATGATGATGATGTCGGGATGGGT 1075
 QY 236 LeuValLeuMetAlaSerYrTrpTrpIleAsnYsAsnValLeuThrAspProArgPhe 255
 DB 1076 CTGCGCAATCTGTTCTCTACTGAGGTGAATGA----- 1111
 QY 256 TyrAsnProGluGluMetGlnIleYsGlyValAlaYsProYsMetAsnMetIys 275
 DB 1112 TATGTGCCCCCTCCCAACCCGTAGCAAGAGAGAGAGAAACCAACAGATGGAAACATG 1171
 QY 276 AspSerPheLeuYrLeuAspArgSerProYrIleLeuLeuLeuThrLeuValIle 295
 DB 1172 GAGAGCTTGAAGTTCTTGAGTGTATCACTACATATGAGGATCTTGACTTGTGTGT 1231
 QY 296 AlaYrGlyIleCysIleAsnLeuIleGluValIleThrPlySerGlnLeuIleGln 315
 DB 1232 GCATATGGAATTAATATCACTTGAAGTCACTAGGAATCAAAAGCTAAAGCTCAG 1291
 QY 316 TyrProAsnMetAsnAspYrSerGluPheMetGlyAsnPheSerPheTrpThrGlyVal 335
 DB 1292 TTCCTTACCCGGAAGAAATACTCAGCATTTATGGCCACTTCTCAACCTGCACAGATTT 1351
 QY 336 ValSerValLeuIleMetLeuPheValGlyIleAsnValIleArgYsPheGlyYrPhe 355
 DB 1352 GCAACATTCACAAATGATGCTT---CTAAGCAATACGTTTAAAGATGATGAGGGA 1408

QY 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
 DB 1409 GTACCTGCAGAAATGATACACACACCGTTCTGCTATTGACCGGTGCTCTTCTCTG 1468
 QY 376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
 DB 1469 ATACTGTTGGTGGCCATTCGACCATGCTGTTGCCAAGCTGGTATGACACCGCTACTC 1528
 QY 396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerIysSerThrIleYrAlaLeu 415
 DB 1529 GCAGCACTGACCTGGTGGCTCCCAATATCTTCGCAAGAGTGCACAGTACAGCTTG 1588
 QY 416 PheAspSerThrIysGluMetAlaYrIleProLeuAspGlnGluIleValIleGly 435
 DB 1589 TTGATCTCTGCAGAAAGATGGCTTAATCTTCCATTGATGAGACACACCAAGTTAAAGC 1648
 QY 436 LysAlaAlaIleAspValValAlaIleArgPheGlyIysSerGlyAlaLeuIleGln 455
 DB 1649 AAAGCTCAATTAATGATGCTGTCGCAACCATTTGGGAAATCAGCGGTCTTAATCAG 1708
 QY 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProYrLeuAlaValIle 475
 DB 1709 CAGTTCATGATCCTTACATTCGGCTCACTCGCCAAATTCACACCTTACCTTGAGATC 1768
 QY 476 LeuLeuPheIleIleAlaIleThrPheValSerAlaThrIysLeu-----Asn 491
 DB 1769 CTGCTCGTATGACTGATGCTGATGCTTTCGACGACTTAAATCGCTTGAGGACAGATTAA 1828
 QY 492 LysLeuPheLeuAlaGlnSerAlaLeuIleGlnGlnIle----- 504
 DB 1829 ACTCTATGCTGAGAAAGCTTGAAGAGGAATGAGAGAGCTTCATCAGTCAAGATT 1888
 QY 505 -----ValAlaGlnGluAspSerAlaProAla 513
 DB 1889 CCTGTTGATCTCAGAGAGAT---GCCGCATCA 1918

RESULT 7
 AAF61374
 ID AAF61374 standard; cDNA; 1896 BP.
 AC AAF61374;
 XX
 DT 25-MAY-2001 (first entry)
 XX
 DE Potato chloroplast ATP/ADP translocator cDNA.
 XX
 KW ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;
 KW adenosine diphosphate; animal feed; energy; biosynthesis; potato;
 KW nutritional value; ss.
 XX
 OS Solanum tuberosum.
 XX
 PN W0200120009-A1.
 XX
 PD 22-MAR-2001.
 XX
 PF 05-AUG-2000; 2000WO-EP07625.
 XX
 PR 15-SEP-1999; 99DE-1044212.
 XX
 PA (BADI) BASF AG.
 XX
 PI Reindl A, Geigenberger FJ, Neuhaus H, Graeve-Kampfenkel K;
 PI Mehlmann T, Tjaden J;
 XX
 DR WPI; 2001-244814/25.
 XX
 PT Transformed plant with an altered content of essential amino acids and
 PT having improved nutritional quality, for use as an animal feed.
 PT comprises a modification in the ATP/ADP translocator gene -
 XX
 PS Example 4; Fig 2; 32pp; German.
 XX

This invention describes a novel transformed plant (A), and its descendants, in which the regulatory sequences and/or the gene copy number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene (1) has been altered so that, compared with the wild-type plant, it has a different content of one or more amino acids (aa). The invention also describes (1) (1), for use in plants, having an approximately 1.7 kb sequence (S1), given in the specification, from *Arabidopsis thaliana*, EMBL Z49227; (2) a genetic construct (GC) containing the (1) and operably linked regulatory sequences; (3) a vector containing the (1) or GC; and (4) seeds, tissues, cells and replicative materials of (A). (A) are useful for human or animal feeding, also (including their cells, tissues and extracts) in agriculture, the animal feed and pharmaceutical industries, and the health service. Altering the activity of (1) increases the amount of ATP in chloroplasts and the amount of energy of available for biosynthesis. Specifically (A) have increased contents of one or more essential aa, so are of greater nutritional value than wild-type plants.

Sequence 1896 BP; 473 A; 369 C; 445 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5.3e-132 Length: 1896
Score: 1306.00 Matches: 259
Percent Similarity: 69.34% Conservative: 87
Best Local Similarity: 51.90% Mismatches: 141
Query Match: 49.66% Indels: 13
DB: 22 Gaps: 3

US-09-869-433-2 (1-515) x AAF61374 (1-1896)

16 LeuTPProIleHisThRiEgIuLeuYslyValLeuProMetPheLeuMetPhe 35
1308 TTAATGGG-ATTGAACCTGTGACCTTAAGAAATATATACCACTGGGCGCATGTTCTT 366
36 CysIleThrPheAspTyrThrValLeuArgAspThrIleAspThrLeuIleValGlyAla 55
367 TGTATTCGTGTTAATTATACAACTCTTAGGATACCTAGATGCTGTGTTGTAACAGCT 426
56 ProGlySerGlyAlaGluAlaIleProPheIleLeuPheLeuValValProCysAla 75
427 AAAGGTCACAGTGTGAGATTAATCCCTTCTTGAACCTGGGTAATTCCTTAGTGGCT 486
76 IleIlePheMetLeuIleTyrAlaIleLeuSerAsnIleLeuSerGlyAlaIleLeuPhe 95
487 ATTGATTCATGCTTTTGTACACAAAGTGGCTAATGCTGTGTCAAGAGGCTCTTTT 546
96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
547 TATACGTATATACCTCTTTATGTCATCTTGGGCGCTTGTGTTGTTGATCTCT 606
116 LeuArgAspValLeuHisProThrGluPheAlaArgGluAlaIleLeuProPro 135
607 CTTAGCAATTACTTACCCCTACAGCTTGTGTGTAAGCTTCTCAATACCTTGGTCA 666
136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTyrPheAlaIlePheTyrValLeu 155
667 AGATTCTTGAACCAATGCTATTTCTGAGATCTGAGATTTCTGCTTTCTATGTCATG 726
156 AlaGluLeuTyrGlySerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIle 175
727 GCTGAGCTTGTGGGAAAGTGTGTGTGCTTCACTTGTGGGCTTGTGTAATCAGATC 786
176 ThrIleIleHisGlyAlaIleValArgPheThrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
787 ACGACTGTGATGAGGCTTAAGATTTATCTTGTGTGAGCTTGTGGGCAAGATGTGCT 846
196 LeuLeuAlaSerGlyAlaGlyAlaIleValTyrAlaSerIleLeuArgAlaSerValSerGlu 215
847 CTATATTTCTCTGTGCGACAGTGAAGTATTTCTTACTTGTGAGAGCTCTTTAGTCTCT 906
216 GlyValAlaProTyrGlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
907 GAGATTGATGTGGGCTATCTCCCTGAAGAAATGATGATGATGTTGATGATGATGCT 966

236 LeuValLeuMetAlaSerTyrTyrTyrIleAsnIleValLeuThrAspProArgPhe 255
967 GGGGCAATCTGTTCTTTTACTGAGTGGTGAATGAATGTGCTCCCAACTGCT--- 1023
256 TyrAsnProGluGluMetGlnIleGlyAlaIlePheProIleMetAsnMetIle 275
1024 -----ACGATG 1062
276 AspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuValIle 295
1063 GAGAGCTTGAAGTCTTGGTCTCTTCAAAATATATCAGGATCTTGCACATTTGGTGTGA 1122
296 AlaTyrGlyIleCysIleAsnIleLeuIleGluValThrThrPheIleValSerGlnLeuGln 315
1123 GCATATGCACTTACTATACCACTTGTGAAAGTTTCAAGAGATGTAAGCTCAAGCTCAG 1182
316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTyrThrIleVal 335
1183 TTCCCAAGCCCAATGAATATCTCTCATTCATGGGTGACCTTCACTGCTACTGAGATA 1242
336 ValSerValLeuIleMetLeuPheValGlyAlaValIleArgIlePheGlyTyrPheLeu 355
1243 GCAACTTTCACAAATGATGTG---TTAGTCAATGATTTTTCAGACAGTATGGGTGGGA 1299
356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
1300 GCAGACGCCAAGATTAACCACTTACAGTCTTGTCTCTTACCGAGTGTGTTCTTCTCCCTG 1359
376 ValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMet 395
1360 CTTTGTTTGGGCGACCTCTAGCACTTACTCTCGAAGTTTGGATGATGACTCTCTCTTA 1419
396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerIleSerThrIleTyrAlaLeu 415
1420 GCAGCTTCTATGTGGGTGGAATGCAGAACATTTTCAGTAAAGTGCAGAAATATAGTTTG 1479
416 PheAspSerThrIleGluMetAlaTyrIleProLeuAspGlnGluIleValValGly 435
1480 TTTACCCCTGCAGAAAGATGCTCTTACCTCTTGTGATGAGACACCAAGTTAAAGCG 1539
436 LysAlaAlaIleAspValValAlaAlaArgPheGlyIleSerGlyAlaIleGln 455
1540 AAGGCAACATTCATGTTGTCTGCATTCACCTGGGAAAGCTGGAGAGCTTGTATCAAA 1599
456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrIleAlaValIle 475
1600 CAGTTCAATGATTTGACTTTGTGCTTGTGCTTGCAGCTGCAGACCTTCTTGGCGGTGTG 1659
476 LeuPheIleIleAlaIleTyrPheValSerAlaThrIleLeuAsnIleLeuPheLeu 495
1660 CTCTTATGATTTGTTCTTGCATGTGTTGGAGACACCAAGCTTGTGATGACAGCTTC--- 1716
496 AlaGlnSerAlaLeuIleGluGlnIleValAlaGlnIleAspSerAlaProAlaSer 514
1717 -----ACTCAATTAAGCAAGAAAGATCTTGAAGAAATGAGAGAGCATCG 1767

RESULT 8
AAF61373 standard; cDNA; 1770 BP.
ID AAF61373;
AC AAF61373;
DT 25-MAY-2001 (first entry)
XX A. thaliana chloroplast ATP/ADP translocator cDNA.
XX ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;
XX adenosine diphosphate; animal feed; energy; biosynthesis;
XX nutritional value; ss.
XX Arabidopsis thaliana.
XX OS

PN W0200120009-A1.
 XX 22-MAR-2001.
 XX 05-AUG-2000; 2000WO-EP07625.
 XX 15-SEP-1999; 99DE-1044212.
 XX (BADI) BASF AG.
 PI Reindl A, Geigenberger PL, Neuhaus H, Graeve-Kampfenkel K;
 PI Moehlmann T, Tjaden J;
 DR WPI, 2001-244814/25.
 XX Transformed plant with an altered content of essential amino acids and
 PT having improved nutritional quality, for use as an animal feed,
 PT comprises a modification in the ATP/ADP translocator gene -
 XX Claim 6; Fig 1; 32pp; German.
 XX This invention describes a novel transformed plant (A), and its
 CC descendants, in which the regulatory sequences and/or the gene copy
 CC number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene
 CC (I) has been altered so that, compared with the wild-type plant, it
 CC has a different content of one or more amino acids (aa). The invention
 CC also describes (1) (1), for use in plants, having an approximately 1.7
 CC kb sequence (S1), given in the specification, from *Arabidopsis thaliana*,
 CC EMBL Z49227; (2) a genetic construct (GC) containing the (I) and operably
 CC linked regulatory sequences; (3) a vector containing the (I) or GC; and
 CC (4) seeds, tissues, cells and replicative materials of (A). (A) are
 CC useful for human or animal feeding, also (including their cells, tissues
 CC and extracts) in agriculture, the animal feed and pharmaceutical
 CC industries, and the health service. Altering the activity of (I)
 CC increases the amount of ATP in chloroplasts and the amount of energy
 CC available for biosynthesis. Specifically (A) have increased contents of
 CC one or more essential aa, so are of greater nutritional value than
 CC wild-type plants.

XX Sequence 1770 BP; 441 A; 383 C; 431 G; 515 T; 0 other;

Alignment Scores:

Pred. No.: 4,97e-127 Length: 1770
 Score: 1260.00 Matches: 262
 Percent Similarity: 69.22% Conservative: 82
 Best Local Similarity: 52.72% Mismatches: 136
 Query Match: 47.91% Indels: 19
 DB: 22 Gaps: 4

US-09-869-433-2 (1-515) x AAF61373 (1-1770)

QY 16 LeuTrpProIleHisThrHisGluLeuLysValLeuProMetPheLeuMetPhe 35
 Db 310 ATTTTCGGTGGAGAGTTGCAACCTTGAAGAAAGATTATCCCTTATGATGATGTTCTTT 369
 QY 36 CysIleThrPheAsnIleThrValLeuArgAspThrLysAspThrLeuIleValGlyAla 55
 Db 370 TGTATTCCTTTCATACATACATTCGAGGATACAAAGATGTCCTTGATGTCGCGCG 429
 QY 56 ProGlySerGlyAlaGluAlaIleProPheIleLysPheTrpLeuValProGlyAla 75
 Db 430 AAGAGAGTTCCTGCGAGATTATACCTTCTTGAAGACTGGGTAAATCTTCCTTAAGGCC 489
 QY 76 IleIlePheMetLeuIleLeuValLysLeuSerAsnIleLeuSerLysGlnAlaLeuPhe 95
 Db 490 ATTGGATTATGCTCTCTACACTTAACCTCTCCATGTTCTCTCCAAAGAGCTCTGTTT 549
 QY 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
 Db 550 TACACTGTTATGTCCTTTCATCATCACTTGGGGCTTGGTTTGCTACATGACCT 609
 QY 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135

Db 610 CTCAGCAATATATATTCACCCGGAAGCTCTGCAGATAGACTCTTACCAACCCCTGGCCCA 669
 QY 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeu 155
 Db 670 AGATTCAATGGGTCTCTATTCGCAATATTCGGAGTTTGGAGTTTCTGTTTATGATTAAG 729
 QY 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
 Db 730 GCTAGCTTTGGGGTAGTGTTGGTGTCTCACTTCTTCTGGGCTTGTCTTACATCAGATC 789
 QY 176 ThrLysIleHisGluAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 Db 790 ACAACTGGATGATGAAGCCAAAGAAATTCATCTTGTTCGACATTTGAGCCCAAGTTGCA 849
 QY 196 LeuLeuAlaSerGlyAlaGalaIleValITrPAlaSerLysLeuAlaGalaSerValSerGlu 215
 Db 850 CTGATTTTCTCAGAGAGAACCGTAATAATCTTCTTAACCTTGAGAAAGAAATCTTGCTCT 909
 QY 216 GlyValAspProTrrpGly-IleSerLeuArgLeuLeuMetAlaMetThrIleValSerGly 235
 Db 910 GGAGTTGAC-----GGCAGTTTGTGGAAGCCATGATGAGCATTTGGTGGAGATGGG 963
 QY 235 YLeuValLeuMetAlaSerTyrTrpTrpIle-AsnLysAsnValLeuThrAspProArgP 255
 Db 964 ACTGCACTTGTCTCTC--TATTGTGGTGGATAGATATGTTCTCTTCCCAACCGCTA 1021
 QY 255 heTyrAsnProGluGluMetGlnLysGlyLysGlyAlaLysProLysMetAsnMetL 275
 Db 1022 GCAAGAAACAAAGAGAG-----AAACCGAAGATGGGAACGA 1057
 QY 275 YAspSerPheLeuTyrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValI 295
 Db 1058 TGGAAGCTTGAAGCTTGTGTATCATCATCATCATATAGATCTTCTACTTATGGG 1117
 QY 295 LealTyrGlyIleCysIleAsnLeuIleGluValIThrTrpLysSerGlnLeuLysLeuG 315
 Db 1118 TGGCATACGGATATAGATATCATCTTGTGGAAGCATCATGAAATCAAGCTTAAAGTTC 1177
 QY 315 InTyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyV 335
 Db 1178 AGTCCCTGAGCCCAAGATGAGTACATCATTTATGGAGCATTCCTCAACCTGCAGCGGTG 1237
 QY 335 aValaSerValLeuIleMetLeuPheValGlyGlyAsnValIleArgLysPheGlyTrpL 355
 Db 1238 TTGCAACATTCACATGATGCTT--CTCAGCCCAATACGTAATGATGATGACCGGTAC 1294
 QY 355 euthrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAla 375
 Db 1295 GAGTAGCTGCAAGATCAACCCCACTGTCTGCTATTTGACCTGTGCTTCTTCTCTC 1354
 QY 375 euValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeu 395
 Db 1355 TAATATGTTTGGGGCCCATTCGCACACACTTGTTCGAAGCTTGGTATGACACGGTAC 1414
 QY 395 etLeuAlaValAlaValGlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAla 415
 Db 1415 TTGAGCTGTGTATGTGGTGGTCCCTTCAGAAATATCTTCAGAAAGATGCCAAGTACAGCT 1474
 QY 415 euPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGluLysValIlySG 435
 Db 1475 TGTTCGACCTTCGAAAGAAATGSCCTTATTCCTCATTTGATGATGACGACCAAGGTTAAG 1534
 QY 435 IlyLysAlaAlaIleAspValAlaAlaIleAspPheGlyLysSerGlyValAlaLeuIleG 455
 Db 1535 GCAAGCTGCAATGATGATGATGATGCAACCATATGAGAAATACAGGGGAGCTTTAATAC 1594
 QY 455 InGlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValI 475
 Db 1595 AGCAGTTTCAATATCTTATCTTGTGATCAGGAAATTCACGCCGTATCTTGAAGTGA 1654
 QY 475 leLeuLeuPheIleIleAlaIleTrpLeuValSerAlaThrLysLeu-----A 491
 Db 1655 TCTTGTGTTATTTGTACCTGCGGTGTTAGCTGACGCTAAGCTCTGAGGAGCAAGTTCA 1714

QY 491 enlyLeuPheLeuAgiInserAlaLeuYgluIngluValAla 506
 DB 1715 ACAGCTTGCGCTCTGAAGAGAGCTTGAG-AGAGAAATGAGAGAGCT 1760

RESULT 9
 AAF84486
 ID AAF84486 standard; DNA; 1823 BP.
 XX AAF84486;
 AC AAF84486;
 DT 02-JUL-2001 (first entry)
 XX
 DE Chlamydia pneumoniae Npt2cp (ADP/ATP translocase) gene.
 XX
 KW Npt2cp; ATP/ADP translocase; respiratory tract infection; pneumonia;
 KW bronchitis; sinusitis; atherosclerosis; asthma; vaccine; antimicrobial;
 KW antibacterial; diagnosis; ds.
 XX
 OS Chlamydia pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 101..1723
 FT /tag= a
 FT /product= "Chlamydia pneumoniae Npt2cp protein"
 FT /function= "ADP/ATP translocase"
 XX
 FN WO200121803-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 15-SEP-2000; 2000WO-CA01087.
 XX
 PR 17-SEP-1999; 99US-0154326.
 XX
 PA (AVENT) AVENTIS PASTEUR LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX
 DR WPI, 2001-316102/33.
 DR P-PSDB; AAB73452.
 XX
 PR New Npt2cp (ADP/ATP translocase) polypeptides and nucleic acids
 PT encoding the polypeptides useful for treating, preventing or diagnosing
 PT Chlamydia infections, particularly infections caused by Chlamydia
 PT pneumoniae -
 XX
 PS Claim 2; Fig 1; 79pp; English.
 XX
 CC This sequence represents the Chlamydia pneumoniae Npt2cp (ADP/ATP
 CC translocase) gene. Chlamydia pneumoniae Npt2cp nucleotides and
 CC proteins may be used for preventing, treating or diagnosing
 CC Chlamydia infections, particularly those caused by Chlamydia
 CC pneumoniae. Chlamydia pneumoniae can be responsible for both upper and
 CC lower respiratory tract infections. It is the third most common cause of
 CC community acquired pneumonia, and can also cause bronchitis and
 CC sinusitis. It is also linked to diseases other than respiratory tract
 CC infections, being associated with atherosclerosis and asthma.
 CC Npt2cp nucleotides may be used for the recombinant production of the
 CC protein, as genetic vaccines, in the construction of vaccine vectors
 CC such as poxviruses, and in the construction of attenuated Chlamydia
 CC strains which can over-express an Npt2cp nucleotide, or which can
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used
 CC in vaccine compositions, and may also be used as diagnostic reagents for
 CC detecting the presence of anti-Chlamydia antibodies in a sample.
 CC
 XX
 SQ Sequence 1823 BP; 450 A; 359 C; 396 G; 618 T; 0 other;

Alignment Scores:
 Pred. No.: 1,93e-101 Length: 1823
 Score: 1025.50 Matches: 214
 Percent Similarity: 59.12% Conservative: 107
 Best Local Similarity: 39.41% Mismatches: 185

Query Match: 38.99% Indels: 37
 DB: 22 Gaps: 9
 US-09-869-433-2 (1-515) x AAF84486 (1-1823)

QY 1 MetThrIySerHrgIuGluIyPProPheGlyLysLeuArgSerPheLeuTyrProIleHis 20
 DB 101 ATGCAGTCATCAGAGAAAGCAACCTTTTCAAGGCTGGCGGCATATCTTGTCTATTTAT 160

QY 21 ThrHISGluLeuYblyValLeuPProPheLeuMetPhePheCysIleThrPheAsn 40
 DB 161 AAATCAGAAATTTCTAGTTTGTTCCACTATTTCTACTAGGCTTTTTCGTGGCTTTAAC 220

QY 41 TyrThrValLeuArgSerThrIyAspThrLeuIleValAlaIleProGlySerGlyAla 60
 DB 221 TACTGCCCTGCTGAAACATATAAGATATCTGTGCTATTCGCTTCAGATCGCTGGGCA 280

QY 61 GluAlaIleProPheIleLysPheTyrPheValValProCysAlaIleIlePheMetLeu 80
 DB 281 GAAGTGATTCCTTCCTTAAGCTTTGGGGAATTCGCCGGAGCTGTATTATTGTTACTATG 340

QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerIySglAlaLeuPheTyrAlaValGlyThr 100
 DB 341 GCTCATGGGTGGTAGGAGAGCGGATCTCGGAGATACCGTTTATTGCTTCATGCGCC 400

QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 DB 401 GCATTCCTGTGTTTCTTCTCTGCTGTGTATCTTATCTGTGAGGGATAGCTTG 460

QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 DB 461 CATCTCAACTCTCTCGCTGATTAATTAACAAGGCTCTTCCCTCAAGAGCTTCGTGGTTT 520

QY 141 ValAlaIleLeuArgAsnTyrThrPheAlaAlaPheTyrValLeuAlaGluLeuTyrGly 160
 DB 521 ATGTGATGATGTCCTTACTGAGATGATTAATTAATTAATTAATTAATTAATTAATTA 580

QY 161 SerValMetLeuSerLeuMetPheTyrGlyPheAlaAsnGluIleThrIyLysGlu 180
 DB 581 TCGGTTGCTCTTGTGATGCTTCTTCTGGGACTACCATGATCATCTACTACATTAATCGAA 640

QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 DB 641 GCGGCGCGTTTATTCGCTTATCAATACAGATTAAATCTCTCAATATGCGCAGGA 700

QY 201 ArgAlaIleValTyrPalaSerIyLeuArg---AlaSerValSerGluGlyValAspPro 219
 DB 701 GAAATCTCTATTGATGGGGAACAAACATTTGTCCTACTCTTTCATGTGATTC 760

QY 220 Tyr---GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeu 238
 DB 761 TGGACTCTGTGAATGCTCACTTACCATGCTGATCTACT---TGTTCGTTTAATTAATG 817

QY 239 MetAlaSerTyrTyrTyrIleAsnIyAsnVal-----LeuThrAspProArgPheTyr 256
 DB 818 ATC-----TGGCTATATAGGGGAGTTGATCATTTGACTTATGATATCTGCATC 865

QY 257 AsnProGluIuMetGlnIySglYblySglYAla----- 268
 DB 866 CCTTCATCTAGACCTGTCTTGGCAGAAAGAGGACACTACTGCTAAATCTAAAGAAAG 925

QY 269 ---LysProIySmetAsnMetIyAspSerPheLeuTyrIleuAspArgSerProTyrIle 287
 DB 926 AAAAACCCTTAAGCAAGCAAGCTTAAGACCTTTCTTACACTCATTCAGTCTCGTATTTA 985

QY 288 LeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCysIleLeuLeuIleGluValThr 307
 DB 986 TTAGGCTCGCTATATATGTCATCTATATATTTGATGATTCATATTTGAGAGTGT 1045

QY 308 TyrIySerGlnLeuIySLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGly 327
 DB 1046 TGGAAAGATCAAGTTTACAGATTTTACAGTTTCACTGATTAATTAATTAATTAATTAAT 1105

Oy	338	Asn	Phe	Ser	Phe	Thr	Gly	Val	Val	Ser	Val	Ileu	Ile	Met	Leu	Phe	Val	Gly	Val	Asn	347		
Db	1106	AGA	TAC	TAC	TAC	CCCT	CAT	TGG	CGG	CGT	CTT	CTG	TAT	TGA	CAG	CTG	TAC	CTT	TAC	CGA	CAG	1165	
Oy	348	Val	Ile	Arg	Val	Phe	Gly	Thr	Phe	Thr	Gly	Ala	Leu	Val	Thr	Pro	Val	Met	Val	Ileu	367		
Db	1166	TGT	ATC	CGT	TA	AAT	GAG	GAT	GAC	TG	CTG	TG	TGT	TAT	AG	TCA	CTC	CA	TGG	TAT	GT	1225	
Oy	368	Thr	Gly	Ile	Val	Phe	Phe	Ala	Leu	Val	Ile	---	Phe	Arg	Asn	Gln	Ala	Ser	Gly	Leu	Val	386	
Db	1226	TC	AG	AC	CG	CT	CTT	TTT	TCG	GA	ACT	AT	TTT	TTC	GCT	CGA	AAA	AG	AC	AT	CT	1285	
Oy	387	Ala	Met	Phe	Gly	Thr	Thr	Pro	Met	Leu	Ala	Val	Val	Gly	Ala	Ile	Gln	Asn	Ile		406		
Db	1286	GGA	GTT	CTT	GAA	TAC	AC	CTT	TGG	CTC	TAC	TGC	TGC	TG	ACT	GAC	TGA	GGA	GAT	GCA	AAT	1345	
Oy	407	Leu	Ser	Val	Ser	Thr	Val	Ser	Thr	Ala	Leu	Phe	Asp	Ser	Thr	Val	Ser	Glu	Met	Ala	Val	426	
Db	1346	CT	AT	CCG	GGA	GCA	AA	T	T	T	A	CG	T	T	T	T	AT	CA	CA	GA	AA	1405	
Oy	427	Leu	Ser	Gln	Glu	Gln	Val	Ser	Gly	Val	Ser	Ala	Ala	Ile	Asp	Val	Val	Ala	Ala	Arg	Phe	446	
Db	1406	CTT	TCT	CC	AG	GAT	AA	AA	AT	AT	CA	T	CG	GA	AA	CG	CG	GA	T	GA	T	1465	
Oy	447	Gly	Val	Ser	Ser	Gly	Val	Ala	Leu	Ile	Gln	Gln	Gln	Gly	Leu	Leu	Val	Ile	Cys	Gly	Ser	Ile	466
Db	1466	GGA	AA	G	T	CT	G	A	G	G	C	T	T	T	A	TT	A	CC	A	G	G	C	1525
Oy	467	Ala	Met	Thr	Pro	Tyr	Leu	Ala	Val	Ile	Leu	Leu	Phe	Ile	Ile	Ala	Ile	Thr	Phe	Val	Ser	486	
Db	1526	GCA	AGT	T	T	AA	AC	T	AT	CG	T	AC	CG	CT	TA	CT	T	AT	T	AT	T	1585	
Oy	487	Ala	Thr	Val	Ser	Leu	Asn	Val	Ser	Phe	Leu	Ala	Gln	Ser	Ala	---	---	---	---	---	Leu	Val	501
Db	1586	GTT	G	CG	T	AT	T	AT	CG	T	AA	GA	AT	ACT	AT	CA	CT	GA	CT	GT	AT	1645	
Oy																							

[illegible]

OY	220	Trp---GlyIleSerLeuArgLeuMetAlaMetThrIleValSerGlyLeuValLeu	238
Dd	716845	TGGCACTCGTAAATGTCTCAACTTGACCAAGTCGATCACT---	716789
OY	239	MetalaserTyrrTriPilleasnuysaaenVal-----LeuthrAspproAarghetYr	256
Dd	716788	ATC-----TGCCTATAATAGGGCGGAATCATCATTTGACTATTGTACTTCGCATC	716741
OY	257	AasnProgluImeGlinsgilylsyIsGlyAla-----	268
Dd	716740	CCTCATCTAGACGTGTCTTGGCAGAGAAGGGAGCGCTACTGCTAATCTAAAGAAAAG	716681
OY	269	--LysProLyMetacnaMetlysAspsSerPhelentyrLeuaspAspserProtyrile	287
Dd	716680	AAAAAACCTTAAGCCCAAAGCTGAAACCTTTCTTACACCTCACTTAGAGTCGTATTTA	716621
OY	288	LeuleuLeuthrLeuLeuValIlealatyrgylieCyseIleasnleuileguValThr	307
Dd	716620	TTAGGGCTCGTATATATGTCCTAATCTTAATATTGSGTAGTCATCATTTGCAAGTCGTT	716561
OY	308	TrylpySerGlnleuLeuLysleugIntyrProaenMetasnAsptyrSergluPheMetGly	327
Dd	716560	TGSAAGGATCAAGATTAGCCAGATTACAAGTCTCACGTAAGTAATCAATGGGTATAGAGT	716501
OY	328	AasnPheSerPheTriThnrglyValIleSerValIleuIleMetleuPheValGlyglyaen	347
Dd	716500	AGATATCAACCACTCATTTGGCGTCGTTGTATTAAGCAAGCTGTACTCTTACCGAGACKG	716441
OY	348	ValIlearglyPheGlyTrypleuthrGlyAlaLeuValThrProvalMetValleuLeu	367
Dd	716440	TGTATCCGTAATAGGGAGTAGCACTGCGGTGTTAGTCACTCACTTGGTAATGTTAGTT	716381
OY	368	ThrglyIleValPhePheAlaLeuValIle---PhearAsmGlnAlaSerGlyLeuVal	386
Dd	716380	TCAGACACGCTCTTTTCGGAATATTGTCGCAAAAAGACACATCTCTATTTTTGGG	716321
OY	387	AlametPheglThrThrProleuMetleuAlaValValGlyAlalegInaseniIle	406
Dd	716320	GGAGTCTTGGAAATGACCTCTGGCTTAGCTGCTGACCTGAGCTGAGAGGATGCAAAATGTC	716261
OY	407	LeuSerLysSerThryLyetyrAlaLeuPheAspsertThryLsgluMetAlatyrIlePro	426
Dd	716260	CTATCCGGGGGACGAATATTACGTTCTTGTATCAACCAAGGAATGGCCTTATCCA	716201
OY	427	LeuAspglngluGlnLysValIysGlyLysAlaAlaIleAspValValAlaAlaArgphe	446
Dd	716200	CTTTCCTCAGAGGATAAAAATCATGGGAAAGCCGCGATGTGATGTCGTTCAAGATA	716141
OY	447	GlyLysSerArgGlyAlaLeuIleGlnngnyLeuLeuValIlecysselyserIlegly	466
Dd	716140	GGAAAGTCGGAGGCTCTTTAATTACCAAGAGGCTGCTGTATTCTCTCTGTGTGCA	716081
OY	467	AlametThrProtyrLeuAlaValIleLeuLeuPheIleAlaIleThrLeuValiser	486
Dd	716080	GCAATTTAAAGCTATGCCCTTAAGTCTTCTCATTTATATGTCGTTGGATGGGTT	716021
OY	487	AlathrLysLeuAsnLysLeuPheLeuAlaGlnSerAla-----Leuys	501
Dd	716020	GTTGGCTATATCGTAAGAATATACCTAGAGCTGCTGATGCTGTAGCAACCTTGAAA	715961
OY	502	-----GluGlnGluValAlaglnglu	508
Dd	715960	CAACCTTAAGAACCTTCTCTTCAATCGTACGTAGAACCCAGGAATCTGTAGAACAAAGA	715901
OY	509	AspserAla	511
Dd	715900	GAAATGGCT	715892
RESULT 11			
ABL74862			
ID	ABL74862	standard; cDNA; 284 bp.	
XC	ABL74862;		

14-MAY-2002 (first entry)
 Corn tassell-derived polynucleotide (cdps) SEQ ID NO:4236.
 Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPS; inheritance; characteristic; growth; development; disease resistance; environmental adaptability; quality; yield; molecular marker; multigene trait; plant breeding; corn tassell; gene; ss.
 Zea mays.
 US2001051335-A1.
 13-DEC-2001.
 16-APR-1999; 99US-0294093.
 21-APR-1998; 98US-082567P.
 (IAGL/) LALGUDI R V.
 (ITOL/) ITO L Y.
 (SHER/) SHERMAN B K.
 Lalgudi RV, Ito LY, Sherman BK;
 WPI; 2002-163647/21.
 Novel purified corn tassell-derived polynucleotide useful for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs -
 Claim 1; SEQ ID 4236; 201tp; English.
 The present sequence describes a purified corn tassell-derived polynucleotide sequence (cdps) comprising a nucleic acid sequence selected from those given in ABL70627 to ABL76833. The cdps sequences encode corn tassell-derived polypeptides (CDPS). The cdps sequences (1) can be used for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs. (1) are also useful in the evaluation, and alteration of desired characteristics associated with growth and development, disease resistance, environmental adaptability, quality and yield, and as molecular markers for studying inheritance of multigene traits in a plant breeding program. (1) can be used to produce a tassell-specific profile of gene transcription, a transcript image, to clone regulatory elements for use in transformation vectors, to express a polypeptide, to identify, isolate or extend identical or related corn tassell nucleic acid sequences from DNA libraries, in nucleic acid hybridisation or amplification technologies, as query sequences to determine homology of known sequences, as probe for use in Southern or Northern hybridisation, and to identify the presence of and/or to determine the degree of similarity between two (or more) nucleic acid sequences.

Db 63 GCAGCTCATGTTGGAGCATGACATTTTCAGTAAAGATGCAAAATACAGTCTTT 122
 Qy 417 AsperThrylsGluMetAlaTyrIleProLeuAspGlnGluValIleGlyLys 436
 Db 123 GATCCTTCGNAAGATGCGATACATTCCTTTGGATGAGATGCAAGTAAAGTAAA 182
 Qy 437 AAlaAlaIleAspValIleAlaAlaArpHeGlyLysSerGlyValAlaLeuIleGln 456
 Db 183 GCGGCTATTGATGTTGTGTGCAACCCCTTGAGGAATGTGAGGGTCTGATCCAGCG 242
 Qy 457 GlyLeuLeuValIleCysGlySerIleGlyAlaMetThrPro 470
 Db 243 TTCATGATCCT-GTCAATGCTTCTCTCGCGAANTGCAACCG 283

RESULT 12
 ABL76356
 ID ABL76356 standard; cDNA; 246 BP.
 XX ABL76356;
 AC
 XX 14-MAY-2002 (first entry)
 DT
 XX
 XX Corn tassal-derived polynucleotide (cdps) SEQ ID NO:5730.
 DE
 XX Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;
 KM inheritance; characteristic; growth; development; disease resistance;
 KM environmental adaptability; quality; yield; molecular marker;
 KM multigene trait; plant breeding; corn tassal; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US2001051335-A1.
 PD 13-DEC-2001.
 PF 16-APR-1999; 99US-0294093.
 XX
 PR 21-APR-1998; 98US-082567P.
 XX
 PA (LALG/) LALGUDI R V.
 PA (ITOL/) ITO L Y.
 PA (SHER/) SHERMAN B K.
 XX
 PI Lalgudi RV; Ito LY, Sherman BK;
 XI
 DR WPI; 2002-163647/21.
 XX
 PT Novel purified corn tassal-derived polynucleotide useful for
 PT determining altered gene expression, to recover regulatory elements and
 PT to follow inheritance of desirable characteristics through hybrid
 PT breeding programs -
 PT
 PS Claim 1; SEQ ID 5730; 201pp; English.
 XX
 CC The present sequence describes a purified corn tassal-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassal-derived polypeptides (CDPS). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (I) can be
 CC used to produce a tassal-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassal nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridisation or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the

CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.
 XX
 SQ Sequence 246 BP; 58 A; 49 C; 63 G; 76 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,81e-17 Length: 246
 Score: 242.00 Matches: 45
 Percent Similarity: 74.68% Conservative: 14
 Best Local Similarity: 56.96% Mismatches: 20
 Query Match: 9.20% Indels: 0
 DB: 24 Gaps: 0
 US-09-869-433-2 (1-515) x ABL76356 (1-246)

Qy 348 ValIleArgLysPheGlyThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIleLeu 367
 Db 6 ATCCTCAGAAAGTTGGGTGGGAGTTGACGCTACATCAGCGCTCGAGTCTACTCCTC 65
 Qy 368 ThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAla 387
 Db 66 ACAGAGATTGGGCTTCTCTCACTGATTTTGTGGTGAGCCATTGACTCTTATAGAC 125
 Qy 388 MetPheGlyThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAsnIleLeu 407
 Db 126 AAGTTGGAGTACACCTTTCCTGCGCGAGCTCATGTTGGAGCATGACAGACATTTTC 185
 Qy 408 SerLysSerThryIleValAlaLeuPheAspSerThryIleGluMetAlaTyrIlePro 426
 Db 186 AGTAAAGTGCAGAAATACAGTCTGTTGATCCTTGCAAAAGATGCGATACATTCCT 242

RESULT 13
 ABX86909
 ID ABX86909 standard; cDNA; 280 BP.
 XX
 AC ABX86909;
 XX
 DT 24-APR-2003 (first entry)
 DT
 XX
 XX Corn ear-derived polynucleotide (cpd) #5369.
 DE
 XX Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022;
 KM SATMON023; structural gene; functional gene; regulatory gene;
 KM corn ear-specific profile; gene transcription; gene expression;
 KM hybrid plant; desirable trait expression; plant breeding program;
 KM inheritance; desired characteristic; growth; development;
 KM disease resistance; environmental adaptability; quality; yield;
 KM multigene trait; plant; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US6476212-B1.
 PD 05-NOV-2002.
 PF 14-MAY-1999; 99US-0313294.
 XX
 PR 26-MAY-1998; 98US-086722P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Lalgudi RV; Ito LY, Sherman BK;
 XI
 DR WPI; 2003-208840/20.
 XX
 PT Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development -
 PT
 PS Examples; SEQ ID No 5369; 390pp; English.
 XX
 CC The present invention relates to the isolation of corn ear-derived

QY 283 ArgSerProTyrIle 287
 |||
 DB 292 TCCTCGAGCTATGCG 306

RESULT 15
 ABX84884.
 ID ABX84884 standard; cDNA; 265 BP.
 XX
 AC ABX84884;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE Corn ear-derived polynucleotide (cpd) #3344.
 XX
 XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;
 KM SATMON023 structural gene; functional gene; regulatory gene;
 KM corn ear-specific profile; gene transcription; gene expression;
 KM hybrid plant; desirable trait expression; plant breeding program;
 KM inheritance; desired characteristic; growth; development;
 KM disease resistance; environmental adaptability; quality; yield;
 KM multigene trait; plant; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US6476212-B1.
 XX
 PD 05-NOV-2002.
 XX
 PF 14-MAY-1999; 99US-0313294.
 XX
 PR 26-MAY-1998; 98US-086722P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Laigudi RV; Ito LY; Sherman BK;
 DR WPI; 2003-208840/20.
 XX
 PT Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with
 PT growth, development
 XX
 PS Examples; SEQ ID No 3344; 390pp; English.
 XX
 CC The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,
 CC functional, and regulatory genes of corn ear. The polynucleotides
 CC sequences are useful for detecting cpds in a sample, for producing
 CC a corn ear-specific profile of gene transcription, for detecting
 CC altered gene expression in inbred or hybrid plants, and for screening
 CC several molecules for specific binding to the polynucleotide. The cpds
 CC are useful to identify, isolate, or extend identical or related
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic
 CC acid amplification or hybridisation techniques to follow the
 CC expression of desirable traits through plant breeding programs.
 CC Preferably, the cpds are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,
 CC quality, and yield of corn. The cpds are also useful as molecular
 CC markers for studying inheritance and multigene traits in a plant
 CC breeding program. The cpds are useful for producing purified corn-ear
 CC polypeptides by recombinant techniques. They are also useful in
 CC diagnostic assays to detect or confirm conditions or diseases
 CC associated with abnormal levels of cdp expression. ABX81541-ABX89140
 CC represent corn ear-derived polynucleotides (cpds) of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/psipdsIDEntry.html.
 XX
 SQ Sequence 265 BP; 73 A; 57 C; 60 G; 73 T; 2 other;

Alignment Scores:
 Pred. No.: 3.67e-09 Length: 265
 Score: 168.00 Matches: 34
 Percent Similarity: 76.92% Conservative: 6
 Best Local Similarity: 65.38% Mismatches: 12
 Query Match: 6.39% Indels: 0
 DB: 25 Gaps: 0

US-09-869-433-2 (1-515) x ABX84884 (1-265)

QY 421 GlnMetAlaTyrIleProLeuAspGlnGluIleValIleGlyLysAlaIleAsp 440
 |||||
 DB 106 GAAATGGCATATATCCCTTTCGATGAAGATATGAAGTAAAGGAAGCGACCTGAT 165

QY 441 ValValAlaAlaArgPheGlyLysSerGlyGlyAlaLeuIleGlnGlnGlyLeuVal 460
 |||||
 DB 166 GTTGTTTGCAACCCACCTGGGCAAGTCGGCGCGCACCTTATCCAAACAGTTGATGATCCTG 225

QY 461 IleCysGlySerIleGlyAlaMetThrProTyrIleu 472
 |||||
 DB 226 ACGTTGGNCCCTGGCAAAATTCGACTTCATCCTT 261

Search completed: November 25, 2003, 10:51:53
 Job time : 1796 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:13:32 ; Search time 22 Seconds

(Without alignments)
990.459 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTTEKPEFKLRSFLMFIH.....AQSLKQEVAGDSAPASS 515

Scoring table:

BIOSDME2
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	2617	99.5	524	US-09-198-452A-369	Sequence 369, App
2	1025.5	39.0	551	US-09-198-452A-653	Sequence 653, App
3	157.5	6.0	907	US-09-198-452A-306	Sequence 306, App
4	135	5.1	545	US-09-252-991A-32219	Sequence 32219, A
5	129	4.9	488	US-09-252-991A-17902	Sequence 17902, A
6	122.5	4.7	713	US-09-335-409-11	Sequence 11, Appl
7	122.5	4.7	713	US-09-568-102-11	Sequence 11, Appl
8	122.5	4.7	713	US-09-568-102-11	Sequence 11, Appl
9	122.5	4.7	713	US-09-568-102-11	Sequence 11, Appl
10	122.5	4.7	713	US-09-568-486-11	Sequence 11, Appl
11	122.5	4.7	713	US-09-568-472-11	Sequence 11, Appl
12	122.5	4.7	713	US-09-567-899-11	Sequence 11, Appl
13	120.5	4.6	534	US-09-252-991A-30725	Sequence 30725, A
14	120.5	4.6	1162	US-09-252-991A-32764	Sequence 32764, A
15	119.5	4.5	800	US-09-134-001C-5655	Sequence 5655, Ap
16	119	4.5	940	US-09-328-352-8165	Sequence 8165, Ap
17	115.5	4.4	408	US-09-107-532A-6563	Sequence 6563, Ap
18	115.5	4.4	546	US-09-252-991A-17104	Sequence 17104, A
19	114	4.3	414	US-09-328-352-6961	Sequence 6961, Ap
20	113	4.3	478	US-09-134-001C-4637	Sequence 4637, A
21	113	4.3	578	US-09-252-991A-18288	Sequence 18288, A
22	113	4.3	1095	US-09-112-096-15	Sequence 15, Appl
23	112.5	4.3	419	US-09-328-352-4974	Sequence 4974, Ap
24	111	4.2	360	US-08-597-226-11	Sequence 11, Appl
25	111	4.2	360	US-08-746-682A-11	Sequence 11, Appl
26	111	4.2	770	US-09-252-991A-25267	Sequence 25267, A
27	110	4.2	478	US-09-107-532A-6105	Sequence 6105, Ap

28	109.5	4.2	540	US-09-107-532A-5467	Sequence 5467, Ap
29	109	4.1	478	US-09-328-352-7771	Sequence 7771, Ap
30	108.5	4.1	568	US-09-328-352-6605	Sequence 6605, Ap
31	108.5	4.1	760	US-09-555-313B-2	Sequence 2, Appl
32	108.5	4.1	816	US-09-252-991A-24105	Sequence 24105, A
33	108	4.1	463	US-09-252-991A-29935	Sequence 29935, A
34	107.5	4.1	459	US-09-328-352-5089	Sequence 5089, Ap
35	107	4.1	417	US-09-328-352-4168	Sequence 4168, Ap
36	106	4.0	512	US-09-107-532A-5262	Sequence 5262, Ap
37	105.5	4.0	360	US-09-355-313B-4	Sequence 4, Appl
38	105.5	4.0	380	US-09-555-313B-24	Sequence 24, Appl
39	105.5	4.0	388	US-08-446-822-8	Sequence 8, Appl
40	105.5	4.0	388	US-09-328-352-4168	Sequence 8, Appl
41	105.5	4.0	388	US-09-328-352-4168	Sequence 8, Appl
42	105.5	4.0	388	US-09-328-352-4168	Sequence 8, Appl
43	105	4.0	566	US-09-134-001C-3355	Sequence 3355, Ap
44	105	4.0	566	US-08-750-723A-2	Sequence 2, Appl
45	104.5	4.0	559	US-09-191-275-2	Sequence 2, Appl
46	104	4.0	483	US-09-107-532A-4123	Sequence 4123, Ap
47	103.5	3.9	1160	US-09-328-352-6826	Sequence 6826, Ap
48	103.5	3.9	367	US-09-134-001C-5557	Sequence 5557, Ap
49	103	3.9	443	US-09-328-352-7567	Sequence 7567, Ap
50	103	3.9	503	US-09-252-991A-32663	Sequence 32663, A
51	102.5	3.9	776	US-09-165-396-3	Sequence 3, Appl
52	102.5	3.9	439	US-09-328-352-7206	Sequence 7206, Ap
53	102.5	3.9	901	US-09-134-001C-5389	Sequence 5389, Ap
54	102.5	3.9	934	US-09-252-991A-19515	Sequence 19515, A
55	102	3.9	477	US-09-134-001C-3467	Sequence 3467, Ap
56	101	3.8	385	US-09-328-352-5641	Sequence 5641, Ap
57	101	3.8	512	US-09-328-352-6192	Sequence 6192, Ap
58	100.5	3.8	521	US-09-107-532A-3961	Sequence 3961, Ap
59	100	3.8	234	US-09-107-532A-3787	Sequence 3787, Ap
60	100	3.8	335	US-09-328-352-4577	Sequence 4577, Ap
61	100	3.8	335	US-09-252-991A-25524	Sequence 25524, A
62	100	3.8	413	US-09-134-001C-5503	Sequence 5503, Ap
63	99.5	3.8	527	US-09-252-991A-17545	Sequence 17545, A
64	99	3.8	2005	US-08-836-325-7	Sequence 7, Appl
65	99	3.8	425	US-09-328-352-7964	Sequence 7964, Ap
66	99	3.8	453	US-09-252-991A-17394	Sequence 17394, A
67	98.5	3.7	600	US-09-252-991A-22396	Sequence 22396, A
68	98.5	3.7	598	US-09-134-001C-3948	Sequence 3948, Ap
69	98.5	3.7	831	US-08-853-659A-53	Sequence 53, Appl
70	98.5	3.7	831	US-08-677-734A-11	Sequence 11, Appl
71	98	3.7	831	US-09-097-053-11	Sequence 11, Appl
72	98	3.7	403	US-09-328-352-7791	Sequence 7791, Ap
73	98	3.7	422	US-09-134-001C-4048	Sequence 4048, Ap
74	97	3.7	473	US-09-252-991A-28845	Sequence 28845, A
75	97	3.7	443	US-09-328-352-7069	Sequence 7069, Ap
76	97	3.7	445	US-09-252-991A-33142	Sequence 33142, A
77	96.5	3.7	502	US-09-328-352-6968	Sequence 6968, Ap
78	96.5	3.7	274	US-09-489-847-278	Sequence 278, Ap
79	96.5	3.7	449	US-09-328-352-7512	Sequence 7512, Ap
80	96.5	3.7	482	US-09-134-001C-4309	Sequence 4309, Ap
81	96	3.7	483	US-09-252-991A-21015	Sequence 21015, A
82	96	3.7	418	US-09-252-991A-30786	Sequence 30786, A
83	96	3.7	462	US-09-328-352-6888	Sequence 6888, Ap
84	95.5	3.6	476	US-09-328-352-5027	Sequence 5027, Ap
85	95.5	3.6	407	US-09-252-991A-21511	Sequence 21511, A
86	95.5	3.6	559	US-09-328-352-5924	Sequence 5924, Ap
87	95.5	3.6	585	US-09-328-352-6133	Sequence 6133, Ap
88	95.5	3.6	623	US-09-252-991A-19867	Sequence 19867, A
89	95	3.6	793	US-09-107-532A-5141	Sequence 5141, Ap
90	95	3.6	364	US-09-134-001C-4201	Sequence 4201, Ap
91	95	3.6	420	US-09-107-532A-4006	Sequence 4006, Ap
92	95	3.6	450	US-09-252-991A-25019	Sequence 25019, A
93	95	3.6	451	US-09-328-352-5044	Sequence 5044, Ap
94	95	3.6	466	US-09-328-352-4381	Sequence 4381, Ap
95	95	3.6	812	US-09-252-991A-22327	Sequence 22327, A
96	95	3.6	803	US-08-836-325-8	Sequence 8, Appl
97	94.5	3.6	865	US-09-252-991A-19339	Sequence 19339, A
98	94.5	3.6	242	US-09-489-847-134	Sequence 134, Ap
99	94.5	3.6	321	US-09-688-019-2	Sequence 2, Appl
100	94.5	3.6	395	US-09-328-352-5819	Sequence 5819, Ap
			400	US-09-198-452A-3302	Sequence 3302, Ap

ALIGNMENTS

RESULT 1

US-09-198-452A-369
 ; Sequence 369, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Grifflais, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; TITLE OF INVENTION: and treatment of infection
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198, 452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 369
 ; LENGTH: 524
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 ; US-09-198-452A-369

Query Match 99.5%; Score 2617; DB 4; Length 524;
 Best Local Similarity 99.6%; Pred. No. 6.4e-274;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MTKTEKPKGLKRSFLMFIHTEHLEKVLPMFLMFCITFNTYTLRDTKDTLIYVAPSGA 60
DB 10 MTKTEKPKGLKRSFLMFIHTEHLEKVLPMFLMFCITFNTYTLRDTKDTLIYVAPSGA 69
QY 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFALFPVIVPLRDVL 120
DB 70 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFALFPVIVPLRDVL 129
QY 121 HPTFADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSMVSLMFWGFANEITKHE 180
DB 130 HPTFADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSMVSLMFWGFANEITKHE 189
QY 181 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMAAMTIVSGVLMA 240
DB 190 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMAAMTIVSGVLMA 249
QY 241 SYWINKNVLTDPFRYPNEEMQKKGAKPKKMKDSTLYIDRSYIILLTLVYAGIC 300
DB 250 SYWINKNVLTDPFRYPNEEMQKKGAKPKKMKDSTLYIDRSYIILLTLVYAGIC 309
QY 301 INLEVTWKSQKLQYPMNDYSEPMGNFSFMTGVSVLIMLVGNAVIRKFGMLTGALV 360
DB 310 INLEVTWKSQKLQYPMNDYSEPMGNFSFMTGVSVLIMLVGNAVIRKFGMLTGALV 369
QY 361 TPVAVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSK 420
DB 370 TPVAVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSK 429
QY 421 EMATIPLDQEKVKGAIDVVAARFGSGGALIOQGLVIGSGIGANTPILAVILFII 480
DB 430 EMATIPLDQEKVKGAIDVVAARFGSGGALIOQGLVIGSGIGANTPILAVILFII 489
QY 481 AIWLVSATKLNKFLAOSALKEOEVAQEDSAPAS 515
DB 490 AIWLVSATKLNKFLAOSALKEOEVAQEDSAPAS 524

```

RESULT 2

US-09-198-452A-653
 ; Sequence 653, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Grifflais, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention

TITLE OF INVENTION: and treatment of infection

FILE REFERENCE: 9710-003-999

CURRENT APPLICATION NUMBER: US/09/198, 452A

CURRENT FILING DATE: 1998-11-24

NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 653

LENGTH: 551

TYPE: PRT

ORGANISM: Chlamydia pneumoniae

US-09-198-452A-653

Query Match 39.0%; Score 1025.5; DB 4; Length 551;

Best Local Similarity 39.4%; Pred. No. 6e-102;

Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

```

QY 1 MTKTEKPKGLKRSFLMFIHTEHLEKVLPMFLMFCITFNTYTLRDTKDTLIYVAPSGA 60
DB 12 MOSSEVPFSPRLRAYLCPIYKSEFSKFPVPLFLAFPGFNVCLIKKMKDTLIVIGSDAGA 71
QY 61 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFALFPVIVPLRDVL 120
DB 72 EAIPIKFWLVPCALIFMLIYAKLSNLSKQALFYANGTPELIFALFPVIVPLRDVL 131
QY 121 HPTFADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSMVSLMFWGFANEITKHE 180
DB 132 HPTFADRLQALIPPGILGLVALIRNMTFAFYVLAELMGSMVSLMFWGFANEITKHE 191
QY 181 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMAAMTIVSGVL 238
DB 192 AKRFYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMAAMTIVSGVL 250
QY 239 MASWINKNV--LTDRFNPNEEMQKKGAKPKKMKDSTLYIDRSYIILLTLVYAGIC 287
DB 251 MASWINKNV--LTDRFNPNEEMQKKGAKPKKMKDSTLYIDRSYIILLTLVYAGIC 306
QY 288 LLLTLVYAGICINLEVTWKSQKLQYPMNDYSEPMGNFSFMTGVSVLIMLVGNAVIRK 347
DB 307 LLLTLVYAGICINLEVTWKSQKLQYPMNDYSEPMGNFSFMTGVSVLIMLVGNAVIRK 366
QY 348 VIRKFGMLTGALVTPVAVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONI 406
DB 367 VIRKFGMLTGALVTPVAVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONI 426
QY 407 LSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKY 466
DB 427 LSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKYALPDSKSTKY 486
QY 467 AMTPEYALVILFIIAIVWLVSATKLNKFLAOSA-----LK-----EQEVAQE 508
DB 487 AMTPEYALVILFIIAIVWLVSATKLNKFLAOSA-----LK-----EQEVAQE 546
QY 509 DSA 511
DB 547 EMA 549

```

RESULT 3

US-09-198-452A-306
 ; Sequence 306, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Grifflais, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198, 452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 306
 ; LENGTH: 907
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae

US-09-198-452A-306

Query Match	6.0%;	Score 157.5;	DB 4;	Length 907;
Best Local Similarity	20.4%;	Pred. No. 7.5e-08;		
Matches 94;	Conservative 95;	Mismatches 163;	Indels 109;	Gaps 26

```

QY 59 GAEMIPETKEML--VVECAIIFMILIAKSNISLTKOLF-YAAGTPELIPFALFPYIY 114
Db 52 GSALP--KIYLGSSLLICVASSLLIYNLFQKHISATRALFIPISLSILCNFYIILSSIF 109
QY 115 PLRDVLPHTPEADRLQALIPPGLLGLVAIILNNMTPAAFYVLAELMGSMYS-LMEWGPAN 173
Db 110 -----AIDPP-----RSPGF--FYRIV-IWSLTIILSYTSPFMGFVD 141
QY 174 EITKIHEAKREYALFGIGANISILASGRALYMAASKLASVSEGDPPNIGISRLIMMTIV 233
Db 142 QFENLOQGRHFICFN-----AIIPLGAL--GSGIILASLVTHTIGIILLFTPAALVT 194
QY 234 SGLVLMASWYMINQV--LTDPR--FYNPEEMQKKGAKKAPNMKOSFYILDRSPYL- 288
Db 195 FPIV-----FVYSLSKSLSDHDLFD-----YGHPRPLSKALKLCPYDKTYTYLL 241
QY 289 ----LTLVLVAYGICINLIEVTWKSQKLOYPNNNDY--EFMGNFSFWTGVSVLIML 342
Db 242 CFYELMOLLALIA-----TEFNLYKLIPEIOLFPAKKEBEFLVAHIGKCSLWISLNMCFAL 294
QY 343 FVGNVAVIRKEFWMLTGAALVTPVMVLLTGIVPEFALVIFENQASGLVAMEGTPIMLAAVYGA 402
Db 295 FAYSRIYKRLD-----VNNIILPAIICFUS-----LPLTWTEKTTISIVLA 336
QY 403 IQNILSKSTKYALPDSTKEMAY--IPLDOEQVKVGAALDVVAARFGSGGALLIQGILV 460
Db 337 M--VVRGGVTAALDDNNQLIITIGVP--NKIRNQ--IRLVESFIEPIGMLVMS---L 385
QY 461 ICGSIGAMTPILAVIILFIITAIWVYSATKLNKLELAOSALK 501
Db 386 VC--FLSSQGVFCIIISLITATILVC--LVRSYAAALEK 421

```

```

RESULT 4
US-09-252-991A-32219
; Sequence 32219 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32219
; LENGTH: 545
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32219

```

Query Match 5.1%; Score 135; DB 4; Length 545;
 Best Local Similarity 21.1%; Pred. No. 9, 5e-06;
 Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26

QY	32	LMPECFINRYVADTKDTLTVG	-----AAGSGEALPF--IKRWLVPCALIFMLIYA	83
DB	143	LVFPAFTLNTY-----DRAAGVWQPLIAEKMSTANDYANINPFQVGTATIGTILGGR		196
QY	84	KLSNLSKQALFYAVGTFFLIPALFPVIVYPLRDVLPHTBPADRLQAILPGLIGLVAI		143
DB	197	FIDRVGVRRAEFLAV-----LWLSLTGAGATLS		226

```

QY      144  LRMTFAAFVY-LAEL--WGSVMLSIMRGPANETTKIHEARKPA--LEFGANISILA 198
Db      227  AAGFWCRFLIGLEANEAPACVKTRLMPFG-----RAVATGFEAGTNVCAIN 278
QY      199  SGRAI-----WASKLRASVEGVDPMGISLTLNMTLIVSGLVMAASYMINKVLTDP 253
Db      279  TPALLPLISWVG-----WQAAFLCWSAL--GLVHLV-FMNRN----- 313
QY      254  RPYNDEKQKKG-----AKPKNMKDSFLYLDKSPYLLTLTLVIA----- 296
Db      314  -YNPDEHPKVKOSELEYIQOEBEPATRVPSQILRRGTWAFALVAISITAPVFWFYLY 372
QY      297  -----YGCINLIEVTWKSQJLKQYPNNNDYSEFGNFRSPMTGVVSVLIML----- 342
Db      373  WLPPFLNQQYGGISGVTO-----GIPLLTLWLTADFG 405
QY      343  FVGNVIRKPFMLTGCALTVPVVLTLTGIVFEAL-----VIERNQASGL-VAMFGTTPMLLA 397
Db      406  SVGGIILS--SMLIGRGMFAIRRLSLMLLFFCTIVGVFAANAISGLWIAV-----LAIA 458
QY      398  VVGAIQ-----NILSKSTKVALFDSTKEMAYIPLDQEKVKGAALDVVAARFG-----XS 449
Db      459  LAVGHAQMTANIWS-----LVMDYPRKHLMSVTFG-----FGGMCAL 497
QY      450  GGALIQGLAVICSGIGAMTPYLAIVLFIATIMVSAFKL 490
Db      498  GGMETPQO--IVGGVLTATNNNYAVLFTMIPIMVITALLTWL 535

```

```

RESULT 5
US-09-252-991A-17902
: Sequence 17902, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 17902
: LENGTH: 488
: TYPE: PRF
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-17902

```

Query Match	4.9%; Score 129; DB 4; Length 488;
Best Local Similarity	19.9%; Pred. No. 3.66-05;
Matches	111; Conservative 71; Mismatches 180; Indels 196; Gaps 23;
Qy	71 VPCATL---FMILYAKLSNIISK-----QALFYAVGTP---FLIF 105
Db	3 VTPCRLTAAGSMCVCPASASRVLSRGLPGPTLCEPSPRAPAPLIVGTGCPMERIFFSF 62
Qy	106 FALPFPYIYPLRDVLHPTEFADRLQALLPGLGLVAILNRWTFAPAFVLAE----LWGS 161
Db	63 RALT---PATLM-ITGSGLS-----TTLARLADKQYDGLWVG 98
Qy	162 VMLSLMWGFA-----NBITKHEAKRFYALFGIGANISLLASGRAIYWASKLRASVS 214
Db	99 ALMANAFGVGLVGKVGRLIARVGHIRAYVACGV-----VTAAVLG 141
Qy	215 EGVDPWGISRLMAMTIVSGLVIMASVY---WLNQNVLTDPFRNPEEMQKKKAKX 270
Db	142 HGLFPM---LPAMTIALRMVGLGIMCOYMTIESLN-----EQADASQRGVVF 186
Qy	271 KMKNKSFLLYDRSPYL-----LTLTLVAYAGICINLIEVTWKSQULKQYPNNKD 321

```

Db 187 GGYAASATLGLVGLGOMILVAHPQLGPELLMLVAFCFALCLVPLALTHKIHPLAPLE 246
Qy 322 YSEBGNF--SFWGVSVLIM-----LFVGNVIRK 351
Db 247 PRFPIRVPOSILTYLVSLVSGLVGSGFYGLAPYANQLGPNBOVLVYGACIFAGLWQMP 306
Qy 352 FGMVLTGA-----LVTPVAVLTGVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONI 406
Db 307 LGWISDRDRAMLIRACAILL---CLFALPL-----ALLQOMPLALLALGIAASM 354
Qy 407 LSKSTKVALFDSTKEMAYIPLDQEQVKGAIDV--VAARFGK-SGGALIOQGLVIC 462
Db 355 L---QFTLYPLAFAFSDHETERVSLTMLLVTFGVGACIGPLAAGALMR----- 403
Qy 463 GSICAMPTYLAVILLFIATLVYSATKLNKF----- 494
Db 404 -LFGANMLYAFVASCALILVWRVHEKVSGLHRVDAPLHHVPTPDNNTSSPLVAALDPR 462
Qy 495 LAQSALKEQVAVQEDSAP 512
Db 463 VDEQAVDEQVYDGEEDAP 480

```

RESULT 6

```

US-09-335-409-11
; Sequence 11, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Liqon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-335-409-11

```

```

Query Match 4.7%; Score 122.5; DB 3; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

```

```

Qy 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPLIFPALFPYIYPLRDVLPHTPEAD 127
Db 16 LALLIVTARASGELARLRQPEVLGELFGV-----VLGSPVAGLAPGFRALFOE 67
Qy 128 RLQALPPGLGLGVALIRNWTFAFYVLAELMGSVMLSMFWGFANEITKIHAKREFAL 187
Db 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Qy 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTIVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVAVILIERESMR 161
Qy 240 ASYWMIN--KNVLTDRFYNPEEMQKKGAKPKANMDSFLYIDRSPYIILLTLVAY 297
Db 162 RSYAQVTLAAGVSEVAMVAVMTSSSYGASPALAVARSAL--ASGFLFMYLV--- 215
Qy 298 GICINLIEVTW-----KSQKLQYNNNDYSEFNGNFSFWGVSVLIMLFGVGNVI 349
Db 216 GRRLTHLAMRVADATRVSKQVSL-----VLVLTFLAALT 252
Qy 350 RKFGM--LTGALTVPVWV-----LITGI-----VEPALVIFRNQASGL--V 386
Db 253 QRLGHLPLGAFALGVLINSAPRTNRPLLDGVQTLVAGLPAVVFVLAKGVVDVSQLRTP 312

```

```

Qy 367 AMEGTEPLMLA-----VVVGAIQNLSTKTKALFDSTKEMAYIPLDQEQVKGAID 440
Db 313 AAMGTVALLLATATTAAYVPALGALG-----GLRSEALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALIOQGLVICSGISGAMTPYAVILLFIAT-----IMVSATKLNKFL 495
Db 362 LIVAIVG-----VELGILL-----SNEAYTMAYAVALTVAASPALILMLE----- 401
Qy 496 AQSALKEQVAVQEDSA 511
Db 402 -----KRAPPOEESA 412

```

RESULT 7

```

US-09-568-102-11
; Sequence 11, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Liqon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-568-102-11

```

```

Query Match 4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

```

```

Qy 76 IIFMLIYAKLSNLSKQ-----ALFYAVGTPLIFPALFPYIYPLRDVLPHTPEAD 127
Db 16 LALLIVTARASGELARLRQPEVLGELFGV-----VLGSPVAGLAPGFRALFOE 67
Qy 128 RLQALPPGLGLGVALIRNWTFAFYVLAELMGSVMLSMFWGFANEITKIHAKREFAL 187
Db 68 PAVGVVLSGI-----SWIGALLLLMAGIEVDVGIIRKARPPAL 107
Qy 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTIVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVAVILIERESMR 161
Qy 240 ASYWMIN--KNVLTDRFYNPEEMQKKGAKPKANMDSFLYIDRSPYIILLTLVAY 297
Db 162 RSYAQVTLAAGVSEVAMVAVMTSSSYGASPALAVARSAL--ASGFLFMYLV--- 215
Qy 298 GICINLIEVTW-----KSQKLQYNNNDYSEFNGNFSFWGVSVLIMLFGVGNVI 349
Db 216 GRRLTHLAMRVADATRVSKQVSL-----VLVLTFLAALT 252
Qy 350 RKFGM--LTGALTVPVWV-----LITGI-----VEPALVIFRNQASGL--V 386
Db 253 QRLGHLPLGAFALGVLINSAPRTNRPLLDGVQTLVAGLPAVVFVLAKGVVDVSQLRTP 312
Qy 387 AMEGTEPLMLA-----VVVGAIQNLSTKTKALFDSTKEMAYIPLDQEQVKGAID 440
Db 313 AAMGTVALLLATATTAAYVPALGALG-----GLRSEALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALIOQGLVICSGISGAMTPYAVILLFIAT-----IMVSATKLNKFL 495

```


Db 362 LIVAIVG-----VELGLL-----SNEAYTMVAVALVTVTASPALLIWLE----- 401
Qy 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 8

US-09-567-969-11
; Sequence 11, Application US/09567969
; Patent No. 6353457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;

Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFPFALEPTVYPLRDVLPHTERAD 127
Db 16 LALIVTARASGELARLRQPEVGLFCGV-----VLGPSVGLAGFHRALFOE 67
Qy 128 RLQAILPPLGLGLVAILRNMTFAAFVYLAELWGSVMSLMWGFANETTKHEAKRFYAL 187
Db 68 PAVGVLSGI-----SWIGALLLLMAGIEVDGILRKEARPGAL 107
Qy 188 FGIGANISLASGRAIIVASKLIRASVSGVDPWGISLRLMAMTVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLSGGLGIVSVTAIVSVIAKVLIERESMR 161
Qy 240 ASYWMIN--KNVLTDPREFYNEEMQKKGAKPKNMKDSFLYDRSPYIILLTLVITAY 297
Db 162 RSYAQTIVLAAGVSEVAAWLVAMTSSSYGASPALAVARSALL--ASGFLFPMVLV----- 215
Qy 298 GICINLIEVTW-----KSQKLQYNNNDYSEFMGNFSFMTGVSVLIMLFVGNVVI 349
Db 216 GRRLTHLAMRWVADATRVSKQVSL-----VLVLTFLAALLT 252
Qy 350 RKFGW--LTGALVTPVWV-----LITGI-----VFPALVIFRNQASGL--V 386
Db 253 QRLGHPPLGAFALGVILNSAPRTNRPLLDVGQTLVAGLPAPVFPVLGMRDVDSQLRTP 312
Qy 387 AMFGTTPMLA-----VVVGAIONILSKSTKYALFSTKEMAYIPLDOEQVKGAID 440
Db 313 AAKGTVALLLATATRAKVPALGARLG-----GLRSEBALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALLIOQGLVITGSGIGAMTPYLAIVLFIIA-----IWLVSATKLKLF 495
Db 362 LIVAIVG-----VELGLL-----SNEAYTMVAVALVTVTASPALLIWLE----- 401
Qy 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 9

US-09-568-480-11
; Sequence 11, Application US/09568480
; Patent No. 6353458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-11

Query Match 4.7%; Score 122.5; DB 4; Length 713;

Best Local Similarity 21.2%; Pred. No. 0.00032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFPFALEPTVYPLRDVLPHTERAD 127
Db 16 LALIVTARASGELARLRQPEVGLFCGV-----VLGPSVGLAGFHRALFOE 67
Qy 128 RLQAILPPLGLGLVAILRNMTFAAFVYLAELWGSVMSLMWGFANETTKHEAKRFYAL 187
Db 68 PAVGVLSGI-----SWIGALLLLMAGIEVDGILRKEARPGAL 107
Qy 188 FGIGANISLASGRAIIVASKLIRASVSGVDPWGISLRLMAMTVSGLV-----LM 239
Db 108 SALGAIAPPLAAGA-----FSALVLDRLPLSGGLGIVSVTAIVSVIAKVLIERESMR 161
Qy 240 ASYWMIN--KNVLTDPREFYNEEMQKKGAKPKNMKDSFLYDRSPYIILLTLVITAY 297
Db 162 RSYAQTIVLAAGVSEVAAWLVAMTSSSYGASPALAVARSALL--ASGFLFPMVLV----- 215
Qy 298 GICINLIEVTW-----KSQKLQYNNNDYSEFMGNFSFMTGVSVLIMLFVGNVVI 349
Db 216 GRRLTHLAMRWVADATRVSKQVSL-----VLVLTFLAALLT 252
Qy 350 RKFGW--LTGALVTPVWV-----LITGI-----VFPALVIFRNQASGL--V 386
Db 253 QRLGHPPLGAFALGVILNSAPRTNRPLLDVGQTLVAGLPAPVFPVLGMRDVDSQLRTP 312
Qy 387 AMFGTTPMLA-----VVVGAIONILSKSTKYALFSTKEMAYIPLDOEQVKGAID 440
Db 313 AAKGTVALLLATATRAKVPALGARLG-----GLRSEBALVAVGLNM-----KGTD 361
Qy 441 VVAARFGKSGGALLIOQGLVITGSGIGAMTPYLAIVLFIIA-----IWLVSATKLKLF 495
Db 362 LIVAIVG-----VELGLL-----SNEAYTMVAVALVTVTASPALLIWLE----- 401
Qy 496 AQSALKEQEVNOEDSA 511
Db 402 -----KRAPPTOESBA 412

RESULT 10

US-09-568-486-11
; Sequence 11, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

```

; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: prt
; ORGANISM: Sorangium cellulosum
; US-09-568-486-11

```

Query Match	4.7%	Score 122.5;	DB 4;	length 713;
Best Local Similarity	21.2%	Pred. No. 0.00032;		
Matches 105;	Conservative 67;	Mismatches 165;	Indels 159;	Gaps 21

```

OY 76 IIFMFIYAKSNISLTKO-----ALFYANGTPELIFMALEPTIYIPLRDVLHHTFEAD 127
      : : : : :
Db 16 LALLIVTARASGELLARRLRQPEVGLFGGV-----VLQPSVIGALAPGFHRLALFQE 67

```

```

Qy 128 RLQALIPPGGLGLVALIRNMTFAAEYVLAEIWMGSWMLSTLMEWGFSNEITTKIHEARFVAL 187
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 68 PAVGVLSGI-----SWIGALLLLMAGIEVDVGLIRKEARPGAL 107

```

```
OY      | : : : : : : : : : : : : : : : : |  
188 FGIGANISLIASGRAI VMASKLRVASVSEGD PMWGSILRLMAMTIVSGLV-----LM   239  
        | : : : : : : : : : : : : : : : : |  
Db 108 SALGAIAPPLAAGAA-----FSAVLDRPDLPSGLFLGI VLSVTANSVIAKLIERESMR 161
```

QY 240 ASYMMIN--KNVLTDFPREINPEEEMQSKKAKPKNNMKSSPLTDKPSYLLLTLLTLLVIAV 297

 Db 162 RSYQVTTLAAGVSEVAAMVLVAMTSSSYGASPALAVARSALL--ASGFLLEMLVLY---- 215

Db 216 GRRLTHLMMRWADATRVSKQVSL -----VLVLTFLAALT 252

Db 253 QRLGHPILGAFALGVLLNSAPRINRPLLDGVQTLVGLFAVFFVLVAGMVDVSQLRTP 312

Db 313 AAMGTVALLTATATRAKVPALCARLG-----GIRGSEALVAVGLNM-----KGSTD 361

```

362 LIVAVG-----VELGL-----SNEAYTYAVVALVTYATSPALLITWE-----401

```

```

Db      402 -----KRAPPTQESSA 412
          |
          ||
          ||
          ||
          ||

```

RESULT 11
US-09-568-472-11
: Sequence 11, Application US/09568472
: Patent No. 7,382,810

```

/ GENERAL INFORMATION:
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Lyon, James
/ APPLICANT: Mojica, Isivan

```

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF PROTHIONES

APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern

```

;
;
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472

```

```

; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/353,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 11
; LENGTH: 713
; TYPE: PRF
; ORGANISM: Sorangium cellulosum
; US-09-568-472-11

```

Query Match	4.7%	Score 122.5	DB 4	Length 713
Best Local Similarity	21.2%	Pred. No. 0.00032		
Matches 105	Conservative 67	Mismatches 165	Indels 159	Gaps 21

QY 76 IIMLIYAKLSNLSKQ-----ALYVAGTBPFLIFALFPVVIYPLRDVLPTEPAD 127
 Db 16 IALILVTRASGELARLRQPEVLGELGGV-----VLGSPVVGALAPGFRALPQE 67

```

QY      128  LQALIPGLGIVALLRNNTFAFYVLAELWQSMVLSMFPGANEIRKIHKEKRFYL 187
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      68  PAVGVVLSTG-----SWIGALLLLLMAGIEVDVGILRKEARPGAL 107

```

QY 188 FGGAGNISILASGRAIWASKIPASVSVDGDPNGISLRIMAMIVSGV-----LM 239
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 108 SALGAIAPPLAAGAA-----FSALTLDRPLPSGLFLGIIVLSTVAHSVIAKVLIERESMR 161

QY 240 ASYIWIN--KNVLTIDPRFYNPEEMQKGGKARKPMNMKDSFYLDKRSFYLLLTLLVLAY 29
 ::: ::: ::: :::
 Db 162 RSYAQVTLIAGVSEVAAMVLVMTSSSYGASPALAVARSALL--ASGELLFMVLV---- 215

QY : : : : :
298 G L I N T L E V I M ----- K S Q L C L O I F N N I S D S E F M G N S F W I G V S L P L I F L E V G N A T 347
Db : : : : :
216 G R L T H L A M R W Y A D A T R V S K G Q V S L ----- V L V L F L A A A L T 252

253 QRTGLHPLLGAFAALGVLTINSAPRTNRPLLDGVQTLVAGI.PAPYFVFLGMRVDSQLRTP 312

Db 313 AAGTVALLLATAATAKVPALGARG---GLRSEALVAVGLNN-----KGGID 361

```

362 LIVAIVG-----VELGIL-----SNEATMYVALVIVTASPAALLIMDE-----401

```

Db 402 -----KRAPPTQESA 412

RESULT 12
US-09-567-899-11
; Sequence 11, Application US/09567899
; Patent No. 7,382,747

```

: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Majeed, Iqbal
: APPLICANT:

```

APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goetlach, Joern
TITLE OR INTENTION: GEMS FOR THE BIOSYNTHESIS OF EPIDIOLONES

```

1 FILE REFERENCE: 4-30582A
2 CURRENT APPLICATION NUMBER: US/09/567,899
3 CURRENT FILING DATE: 2000-05-10
4 PRIOR APPLICATION NUMBRR: 09/335,409

```

```

: SEQ ID NO 11
:
: SOFTWARE: Patentin Ver. 2.0
:
: NUMBER OF SEQ ID NOS: 30
:
: PRIOR FILING DATE: 1995-06-17
:
:

```

```

: TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-09-869-11

Query Match      4.7%; Score 122.5; DB 4; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.0032;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy 76 IIFMLIVAKLSNLSKQ-----ALFYAVGTPPLIFALFPYIYILRDVHPTERAD 127
Db 16 LALLVTRASGEARLRQPEVIGELRGV-----VLGFSVGLADGFHALPQOE 67
Qy 128 RLQALPFGGLGLVAILRNTFAAFVYLAELMGSVLSTMFGEFANEITKIHAKRFYAL 187
Db 68 PAVGVLSGI-----SWIGALLMLMAGLEVDVGIIIRKARFAL 107
Qy 188 FGIGANISLASGRAIYVASKLRASVSEGVDPWGISLRLLMAMTIVSGIV-----LM 239
Db 108 SALGANIAPPLAAGAA-----FSALVLDRLPLSGFLGIVSVTVASVIAKVLIERESMR 161
Qy 240 ASYMMIN--KNVLDPRFYNPEEMOKKAKPKKMMKDSFLYIDRSFYILLTLVIAY 297
Db 162 RSYAQVTLAAGVSEVAAVLMVAMTSSSYGASPALVARSALL--ASGFLFMVLYV--- 215
Qy 298 GICINLIEVTW-----KSQKLQYNNNDYSEFMGNFSFMGVSVILMLFVGKAVI 349
Db 216 GRRLTHLMRWVADATRVSKQVSL-----VLVTFPLAALT 252
Qy 350 RKFGH--VTGALVTPVWV-----LITGI-----VEPALVIFRNOASGL--V 386
Db 253 QRLGHPILGAFALGVILNSAPRTNRPLLDGVQTLVAGLFAPVFVVLAMGVDSQLRTP 312
Qy 387 AMFGTPLMLA-----VVGAIONLSKSTKYALFDSTKEMAYIPLDOEKVWKAID 440
Db 313 AAMGTVALLTATVAPAKVVPALGARLG-----GLRGEALVAVGLNM-----KGGTD 361
Qy 441 VVARFGKSGGALIQOGLVIVCGSIGAMTPYIAVILFLIA-----IMVSAIKLKLFL 495
Db 362 LIVAIIVG-----VELGL-----SNEAYTMAVAVLVTVASPLLWLK----- 401
Qy 496 AQSALKEOEVAQEDSA 511
Db 402 -----KRAPPTQESBA 412

RESULT 13
US-09-252-991A-30725
: Sequence 30725, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-07-27
: PRIOR APPLICATION NUMBER: US 60/094,190
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 30725
: LENGTH: 534
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30725

Query Match      4.6%; Score 120.5; DB 4; Length 534;
Best Local Similarity 20.0%; Pred. No. 0.0034;
Matches 97; Conservative 71; Mismatches 145; Indels 171; Gaps 24;

Qy 63 IPFIKFMVLVPCALIFMLIVAKLSNLSKQLEFVAVGTPPLIF--FALPPTIYV-----LR 117
Db 63 IPFIKFMVLVPCALIFMLIVAKLSNLSKQLEFVAVGTPPLIF--FALPPTIYV-----LR 117

```

```

Db 110 LPARRFMNVVLCPLPAIPAVSSFTWVSLSPMVEGLGAILVMTSKPYLVLPVAATLR 169
Qy 118 DVLHPTERADR-----LQAILP-----PGL-----GLVAILRNMTE- 149
Db 170 GIDSLSEASAMLGISROYFLRVTLPLRPTLATGILVALHMLVERGALSTIRYQTF 229
Qy 150 AAFV-----VLAELMGSVMLSMF--WGFANEITKIHAKRFYALFGIGANISL 197
Db 230 TAIQEFELERSNATAMLSVLALCFLLM-----LELMR 267
Qy 198 ASGRAI--VWASKLRASVSEGVDPWGISLRLLMAMTIVSGI--VIMASYMMINKNVLTDP 253
Db 268 GRGLVTRTQGSARABRVRVRFKMPQLQALLATLVVGTSIPLVMGYW----- 317
Qy 254 RFYNPEEMOKKAKPKKMMKDSFLYIDRSFYILLTLVIANG--ITINLIEVTWK 309
Db 318 -----LYEGSSASFPLMEIYST--LTSLSLARGGALLSCLLALPV--- 356
Qy 310 SQLKLOYPNNNDYSEFMGNFSFMGVSVILMLFVGKAVIRKFGMLGALVTPVWVLLTG 369
Db 357 SILVVRTP-----GALARMT-----QRLPYLQAL--PGLVIALS 389
Qy 370 IVEPALVIFRNOASGLVAMFGTTPMLAVVGAIONLSKSTKYALFDSTKEMAYIPLDQ 429
Db 390 LVYFSL-----HYLDATYOTGLLVA-----YALL-----FMRPLQ 421
Qy 430 E---QVYKGAIDVVAARFGKSG--GALLQOGLVIVCGSIGAMTPYIAVILFLIAIWL 484
Db 422 APIRVALEKASPOQEBAARTIGTPLMAFLRITLPIISPAIGA-----GFVLVLDTWKE 476
Qy 485 VSAT 488
Db 477 LTAT 480

```

```

RESULT 14
US-09-252-991A-32764
: Sequence 32764, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-07-27
: PRIOR APPLICATION NUMBER: US 60/094,190
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 32764
: LENGTH: 1162
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32764

Query Match      4.6%; Score 120.5; DB 4; Length 1162;
Best Local Similarity 21.1%; Pred. No. 0.0011;
Matches 81; Conservative 65; Mismatches 136; Indels 101; Gaps 19;

Qy 71 VVPCALIFMLIVAKLSNLSKQLEFVAVG--TPPL-----IFPALPPTIYVPLRDVILHP 122
Db 688 VIGIAVV-LIGYAAWAMWLIR-LIFSSPGDERPSELIMIVGILFTALPLVLF----- 737
Qy 123 TEFADRLQALILPPLLGLVAILRNTFAAFVYLAELMGSVLSTMFGEFANEITKIHAK 182
Db 738 -----VACGCGYITTSKLTDRDLIDTLYL--LML--FYIEAAVR--- 774
Qy 183 RFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISLRLLMAMTIVSGI-----SLRLLM 228
Db 775 -----GLGVAAARLAVARAL--AKRQNAKEGVDSDETIVNEPTLGIHQINQSGSLIR 825

```

QY 229 AMTIVSGLVAMSYWINKVLTDPFYNPEEMOK-----GKKGAKPKMMKDSFLYDR 283
DB 826 LALLIGFIVCL---YVWADLISVSYLDNVTLYQTACTGEAANSVPSLAD----- 875
QY 284 SPYILLTLTLVYAGICIN---LIEVTKSQKLOQYNNNDYSEFMGNESFWTGVSVLI 340
DB 876 --FLAALAAVAVTVALARM,PGILEVLYVQRLTLAQSGSAVATTLISVAISGIGIVSALS 933
QY 341 MCFVGNAVIRKRGVLTGALVTVMVLTGIVFEFALVIFPNQASGLVAMFGTPEMLAAVY 400
DB 934 TLGVSWD---KQOMLVAA-----LSVGLGHGLQATFENFVSGLTILF-ERPVRIGDV 982
QY 401 -----GAIQNILSKSTKYALPD 417
DB 983 TIGNSGTVSRIRIRATITTD 1005

RESULT 15

US-09-134-001C-5655
; Sequence 5655, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO: 5655
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5655

Query Match 4.5%; Score 119.5; DB 4; Length 800;

Best Local Similarity 20.3%; Pred. No. 0.0008;

Matches 108; Conservative 86; Mismatches 148; Indels 191; Gaps 30;

QY 2 TKTTEKPPGKLSPLM-----PI-----HTEHKVLPWFLMF-----F 35
DB 222 TKSQAPF---HIMLPKMAAPTPVSAIYHSATMVA-GIFLFRFPPLIGLSNVYI 275
QY 36 CITF-----NYTVLR-DTKDTL-----IVGAPGSGAE-AIPPIKRW 69
DB 276 TTVFGLITMLFGSLTALRQYDLKGLAVSTISQGMIMTWVGLGGVAAQHTSDELSKFY 335
QY 70 LVVPCAIIFMLIYAKLSNLSKQALFYAV----- 98
DB 336 ILVLPAGGFHL---MNAHVFKCALFMGIGIIDHSGTRDIRLLNGMRKVPKMHIVML 391
QY 99 -----GTPPLIFLFPVIYLRDVLHPTEFADRLQAILRPGV--LGVALLANMTPA 150
DB 392 AALSMAGVPFLNGFLSKEMFL---DSLTKANELDQYGFVLFPVITISIGVLSILTFIYA 447
QY 151 AFYVLAELMGVSLMFWGFAN---ELTKIHEAKRF-----YALFGIGA 192
DB 448 LY-----MIKETFMGNINIEKFKKHHEPMLFSLPAVILMLLPIVIFPVNVEG 497
QY 193 NISLASGRAIYMAKLRASVS---EGVDPW-GISLRILMA-MTVVSGIVLMASVW--- 244
DB 498 NFVILPATRSV---SGIGAEVDAPFPHISQMHGVMLPILISIVIIIGIILALVVMKXV 554
QY 245 ----LNKAVLTD--DRFVNPEMOKGKGAAPKMMKDSFLYDRSPYILLTLTL---V 294
DB 555 THQITKASITDGYRKIRYREFELYS-AKIRALMNNKLN-----YIMITLTFIYAI 605
QY 295 IAYGICINLIEVTKSQKLOQYNNNDYSEFMGNESFWTGVSVLIMFVGGNAVIRKFGW 354

DB 606 VVYG-----YLVGFPPHVHQ--HISBGPLEVILSV-VTLIGISILFIQR 650
QY 355 LTGALVTPVMTLIGIVFEFALVIFPNQASGLVAMFGTPEMLAAVYAGIOMIL 407
DB 651 LT-----MVLNGMIGFAVTLY-----PIAKAPDLALQLVETITTL 690

RESULT 16

US-09-328-352-8165
; Sequence 8165, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO: 8165
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8165

Query Match 4.5%; Score 119; DB 4; Length 940;

Best Local Similarity 20.5%; Pred. No. 0.0011;

Matches 141; Conservative 85; Mismatches 210; Indels 252; Gaps 35;

QY 15 FLWPIHTEHKV--LPMFLMFCITENYVLRDTKD---LIVGAPGSGAEAIPIKRW 69
DB 264 FLTVFSSHFPVINSLTASTILFTILATVFLSKQQAIVATIALQGM-AVAAP----- 317
QY 70 LVP---CALIFMLIYAKLSNI-----LSKQALFYA--VSTPPLIFP--AL 108
DB 318 LVIPQYRDVFLPSYLVINLVAANVFQPKILNQIAFAFMFPGSAIAFYAEPAK 377
QY 109 FPTVYLP-----RDVLHPTEFADRLQAILRPGGLVAILRMWTF 149
DB 378 FDTLDWILMLIALFIMLSVRSYSONISKVSHKQGRPLPLDLVGLIRNPVL-GTTL 436
QY 150 AAFVY---LAEIWSVWL-S---LMFWGFRANETIKHE-----AKRFYAL-FGIGANI 194
DB 437 HAVLVHSTQALTIGAAVLACTVAVLTGW-----IKTHPQLSVLASSFFTLAAVAFALI 491
QY 195 SLASG---RAIYVASKLRASVSEGVDPWGISLKLMAMTIVSGIVLMAS-----YMWIN 246
DB 492 FPLKGAHMTAIGVAAQTALI-----VMGVTERYRLSRXYGIVLIVLSSIALFYQVWAN 546
QY 247 KNAVLT-----DRFVNPEMOKGKGAAPKMMKDSFLYDR----- 283
DB 547 EEPFLTSIYAIKAFISAFILQYNSKEQYFBSASM-FSGIFLCLGNYAGAVGEIWA 605
QY 284 -----SPYILLTLTLVYAGICINL-LIEVTKS-QLKLOQYNNNDYSEFMGNESFWTGV 335
DB 606 MHHHALSPYLMFALIALAIFSAIYHYKLRVOMQSLQIL----- 644
QY 336 VSVLIMFVG---GNVIRKGMLTGALVTPVMV---LITGIVFEFALVIFPNQASGL--- 385
DB 645 ISLILLVLGFAFASQVFTLFPKWDLSQOTFLVSTIISLSLFTMA---OPQSHLGYV 700
QY 386 -----VMPFGTPEMLAAVVGAGIOMILS-KSTKXALF----- 416
DB 701 KVMAGLSMALATAYGVAIFPMPIVALAFVYVVSIMVAYRAHKTTLLHQIIVWCLSLIWL 760
QY 417 -----DSTKEMAYIDP-----DOEKVKGAAIDVVAAR 445
DB 761 LVSVVDHSAEYLAFVLENTDFFSILVFAGLFIITYQAHFDDKSLMEWFKTITLIV- 819
QY 446 FGKSGGALIQGLIV-----ICGSGAMTPYAVLILFTIA----- 481
DB 820 -----GLVFSVVVRGLHHYMATPLMSASITWGTGVQVSLTLLMVLIAFLITTYSSRKM 874

QY 482 ----IMVSAT-----KINKFLAOSA 499
| | | | |
Db 875 IQIMFVGAALGIIVLKLILDLGSA 902

RESULT 17

US-09-107-532A-6563
Sequence 6563, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6563:

SEQUENCE CHARACTERISTICS:

LENGTH: 408 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...408

SEQUENCE DESCRIPTION: SEQ ID NO: 6563:

US-09-107-532A-6563

Query Match 4.4%; Score 115.5; DB 4; Length 408;

Best Local Similarity 20.3%; Pred. No. 0.00078;

Matches 88; Conservative 70; Mismatches 162; Indels 113; Gaps 16;

QY 102 FLIPALPTVYIPRDVHPTPEFADRLQAILPGLLGVALT-----RMT 148

Db 27 YIIVLVMYIVADTILH---ASLSQAGPAGSYIIGTLLARLFMGKQLEFGKFT 82

QY 149 F---AAFYVLAELMGSVLMFLMFWGPFANEITKIHAKRFYALFGIGANISILASGRAIYW 205

Db 83 LAGGAIFY-----LLSTLALVLTPTIGALY-LVRFLNGPGY---TVSTATNAIYT 129

QY 206 ASKRLASVSEGVDPKGISRLMLAMTIVSGVLVMASYMINKVLTDPFPYNEEMQK 265

Db 130 AVTPASKEGEGINYGLSTSLAAVGPFLGMITL----- 162

QY 266 KAKAPPMNKDSFLYIDRSPYITLLTLVIAVGICINLIEVTWKQLKQYNNNDYSR- 324

Db 163 -----LMMTD-FRFLMPSIVLVEFVTIACLVFPVKPISEEQKISRNNINSFIEK 215
| | | | |
QY 325 ---FMGNSEFWTGV--VSYLIMLFGANVIR-----KFGMLTGAIVTPMVLITGVF-- 372
| | | | |
Db 216 KALFITGIAFLMGLSYSSVLSFSSYTKVILHVAAGSFFVVYALVITRPLTRIPDV 275
| | | | |
QY 373 -----FALVIFRNOAGSLVMPGTTPLMLAVVGAIONILSKSTKYALPSTKEMAYI 425
| | | | |
Db 276 KGEQVWMPYSYLF--LTGLFLSVTINSVTILSGLVGL-----GVTFMNGOAVCL 328
| | | | |
QY 426 PLDOEKYKGAIDVVAARFGSGGALIQ-----QGLVIGSIGAMTPYLAVAL 476
| | | | |
Db 329 KVRHRHSIALSTYFIDGLDGLGVGYLMGLKSLTSPRGIVYAG-----VIP 378
| | | | |
QY 477 LFIIMLVSATK 489
| | | | |
Db 379 LICTALYLVNGRK 391
| | | | |

RESULT 18

US-09-252-991A-17104

Sequence 17104, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 17104

LENGTH: 546

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17104

Query Match 4.4%; Score 115.5; DB 4; Length 546;

Best Local Similarity 17.2%; Pred. No. 0.0012;

Matches 104; Conservative 77; Mismatches 178; Indels 247; Gaps 22;

QY 18 PIHTHELKVLPMFLMFCITFENYVLRDTKDTLIVGAPSGAEAIPIPKFWL----- 70

Db 18 PIKTHQVRLNDI-----VFFGSTALLIVLTALILYFDSGALLRIQAMLSHSPGWY 71

QY 71 -VPCAILPMLIVA-----KLSNLSKQALFYAVGTPLIFPALFPVIVY-----P 115

Db 72 YMLAIGAVLIFVAWAFSFRFTLKGHEHEKDPFGYAGWGLFSSGIGISILLYPAASEP 131

QY 116 LRDVHPTPEFADRLQAILPGLLGVALLRMTTFAAFVLAELMGSVLMFWGPFANEI 175

Db 132 IDHLVHP-----PEGVPSGPOARQ-----ALDUTFLHMGJ----- 162

QY 176 TKIHAKRFYALFGIGAN-----ISILASGRAIYASKLASVSEGVDPKGISRL 226

Db 163 ---HGMALYALVGLAVGFARHRQPLALSALYPIIGERVRVKAAGAVDCPQIFVTL 218

QY 227 L-----MAVTYSGVLVMASYMINKVLTDPFPYNEEMQK 252

Db 219 LGLVTLNIGALQVSSGLEVLTGMPHSGKTLIAVLAMSLVATLAAVSGV----- 268

QY 253 PRFVYNEEMQKQKAKPKAMKDSFLYLDSSPYITLLTLVIAVGICINLIEVTWKSQL 312

Db 269 -----EKQIRLSNLN-----IVLSSLLFLVLCGSTLEL-----L 300

QY 313 KQYNNNDY-----SEFMGNFSF-----WTGVSVLIML 342

Db 301 NGFVNLDGDIYDKVLTFFDLVYVTGAGAGKSEEMLGIAMTLFYMAWMISMALPFGMFLAR 360

```

QY      343 FVCGNVIRKFGMLTGLVTPVWMLTGLVFEALVIFRNQASGLVAMFGTLP----- 394
DB      361 ISRGSRVRE-----LVCGVLLTLPDGLTAMLSVGNLSLDLVNHNAGTDLKALAEOP 413
QY      395 -----MLAVVGAIGNILSKSTKYALFDSKEMAYILPDOBKVKGAIDVY 442
DB      414 SMSIVLLBHYPLSKVIG-----LSIPGFLV-----LTPADSG-----SWV 452
QY      443 AARFGSG-----ALIQGLVITCSIGM-----TPYLAVL 476
DB      453 LAMLSRSGELDEBDAPNWLRIIMSAVVTLVITGL-FAGNFMQTVVVLAGLPSAVLI 511
QY      477 LPTIAI 482
DB      512 LFMFGL 517

RESULT 19
US-09-328-352-6961
; Sequence 6961, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6961
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6961

Query Match
Best Local Similarity 4.3%; Score 114; DB 4; Length 414;
Matches 81; Conservative 71; Mismatches 130; Indels 72; Gaps 20;

QY      163 MSLMFWGPFANETTKIHEAKRFYALFGIGANISLASGRAIYVASKLRSVSEGVDPW-G 221
DB      109 IISMSPFLPASPVSMLFIAR---GLQIATGLAVSAIGAIIDPSKLGSLINSIAPMIG 165
QY      222 ISLRILMANTIVS-----GLVL-MASYWINKNVLTDRFPNPEMOGKXGA----KP 270
DB      166 MAVGIFLTSIIQFSAHPLQVFEPLCFLLICELLIS---FLTPETAQR-RSGALASLRP 221
QY      271 KM---NMKDSFLYDRSPYILLTLTLVAVGICINLIEVWKSQKLQYPMNNDYSEPM 326
DB      222 NMAIPQTSALIST--SPINILMNV---SGFFSLM-----PSL--LAKIF 262
QY      327 GNFSFWTGVVSVLIMLFVCGNVIRKFGMLTGLVTPVWMLTGLVFEALVIFRNQASGLV 386
DB      263 HTSSAWLNGIMFWALMSG-----VGILTKRSTFRILLTGTSIALI-----GAI 309
QY      387 AMFGTLPMLAVV--VGAIQNLISKSTKALPDSYSTEMAYILPDOGQKXKGAIDVVA 444
DB      310 VFIALINLNAVVLFGSIITGFGGT--AFMGAISSVMPPLPBER-----AGIMAA 360
QY      445 RRGSGGALIQGLV--ICGSIGAMT---PYLA-VILLFIAIYVSAATKLNK 492
DB      361 FVEASTLASIRAILAGVVGKIGLSTANSYSIFILLSLVALMLTIKPKNK 414

RESULT 20
US-09-134-001C-4637
; Sequence 4637, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4637
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4637

Query Match
Best Local Similarity 4.3%; Score 113; DB 4; Length 478;
Matches 94; Conservative 75; Mismatches 174; Indels 138; Gaps 20;

QY      11 KLRSLMPPIHT---HELRKVLV-----MFLMF-----CITFNVTY----- 43
DB      12 KRRKFLPIVGLIILMALPIKPDALNDQAMFMFALPSTIACITQPTTAVSIGRT 71
QY      44 -----LRDTKTLIVAGPAGGAEAP-----PIKFWLVPCAIIMLYAK---L 85
DB      72 IMILVIGVDTK-TAVQFGNSSIWLIMAFISRGFVGTGLRRITALQFKLGKTLGL 130
QY      86 SNILSKQALFYAVGTPFLIFPA---LPTVITYPLADVL--HPTFRADRLQALPPGLGL 140
DB      131 AVSLGVVDLILAPAPPSNTRARGIMFP-IKSLSESFGSSPRDSER----- 177
QY      141 VALINMTFAFYVLAELMGSVMLSLMFGEFANETTKIHEAKRFYALFG--IGANISLTA 198
DB      178 -----KMGAFILITEPQGNLITAMF-----LTAMGNPILQSLAEKT 215
QY      199 SGRAIYVASKLRSVSEGVDPWGISLRILMANTIVSGVLWASYWINKNVLTDRFPNPE 258
DB      216 AHVQITMMWFEVAALIPG-----LISLIVVPFIY-----KLYPP 250
QY      259 --EENQKGGKAKPKMNKDSFLYDR---SPYILLTLVAVGICINLIEVWKSQKL 313
DB      251 TYKETPNAKWMTLEOLEENGHSIAEKLVGVFTALALMWLGSTINDATITARIALAL 310
QY      314 IQYPMNNDYSEPMGNFSFWTGVV--SVLIMLFVCGNVIRKFGMLTGLVTP-----V 363
DB      311 LLLTGLVAMSDILNLTGAMNTLWVFSVLVLMAEQNLKGFIPWLSKLAQGLNGRSPV 370
QY      364 WLLTGLVFEALVIFRNQASGLVAMF-----GTPPLMAYVGAIGNILSKSTKY 413
DB      371 LVLLILFEYSHYLSASATAHVASAMYALLGVAVASGAPPLFSALMLGFFGNLSTHY 430
QY      414 A 414
DB      431 S 431

RESULT 21
US-09-252-991A-18288
; Sequence 18288, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18288
; LENGTH: 578

```

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18288

```

Query Match 4.3%; Score 113; DB 4; Length 578;
Best Local Similarity 22.4%; Pred. No. 0.0025;
Matches 120; Conservative 65; Mismatches 153; Indels 198; Gaps 33;

QY	MTKIEKSPKGLKRSFLMPFIHTEHLKKVLMPMFECITFNVTYLRDTHQTLVYGARGSGA	60
Db	170 MHRSGARP--RLRS-----KRSVLAVLPIILL--AAAGFTIL--TTEFVIIG-----	210
QY	61 EAIPFIKFWLVVPCAILFEMIYAKTLNIIKSOKALFYAVGTPPELLIFALPEFTVIYPLRDVY	120
Db	211 -----LIFPALADIQVSAQGLVLSLFFAFSAAGFPF-----	243
QY	121 HPTFEPADLQAILPPGLGL-----VALRMWTFAPFYVIAELMSGV-----ML	164
Db	244 -----LTLAALQOVERKRLFLVACL--LIFPANAALAAVADITMTAAVARFVPAL	289
QY	165 SL-MFWGPANETITKIHFA-----KRFYAL--FGI--GANISLASGRAIYWASKRASVSE	215
Db	290 ALPFWMAVASE--TAHLAGPSREGRVAVLPFGVIAATVIGIPIGTLI-----	336
QY	216 GVDPPG--ISLRLMAMTIVSGLYVMASVWYINKVLTDPRAFNPBEMOKKKGAKEXKN	273
Db	337 -ADAWGMRFLAPALALALAKALLLAA--WL-----PRI-----PGRPGIS	374
QY	274 MKDSFLYADRSEFYI--LTLTLVLVAYGICINLIEVWKSQULKQYPMNDYSE-----FM	326
Db	375 LRSQASVL--RQRLVIGHLLSLLVF-----TGMETPYTLADILQRLAGFS	419
QY	327 GNFSFWT-----GVSVLMLFVGVGNVIRKFGMLTALV-----TPWVLLTGIVFPAL	375
Db	420 GSLVGWMTLMGFVAVGLL-----GN-----WGGGLVVRSPGATLLFVLLMALQMLAL	467
QY	376 VIFRQQA---SGLVAMFGTTPMLLVVVVGAIONILSKSTRYALFDSKEMAYIPLDQEK	432
Db	468 VPTLGNAMLLATLAIMGVAQAL--FIVQVVR--VMSARAPRAF-----	509
QY	433 VKGKRAIDVVARFGSGAL-----IQGGLVIGSGISGMW--PYLVAILLEFIIA	481
Db	510 ---ASLNIISACNAIGIGALGWSVIDSGALPDEBVALVCLAMAAVALLMLA	562

```

RESULT 22
US-09-112-096-15
Sequence 15, Application US/09112096
Patent No. 6194152
GENERAL INFORMATION:
APPLICANT: Reiner Laus
APPLICANT: Michael H. Shapiro
APPLICANT: Larissa Tsavaler
TITLE OF INVENTION: Prostate Tumor Polynucleotide and
TITLE OF INVENTION: Antigen Compositions
FILE REFERENCE: 7636-0015.30
CURRENT APPLICATION NUMBER: US/09/112,096
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/056,110
EARLIER FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 29
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 1095
TYPE: prt
ORGANISM: Homo sapiens
US-09-112-096-15

```

Query Match	4.3%	Score 113;	DB 3;	Length 1095;
Best Local Similarity	21.0%;	Pred. No. 0.0064;		
Matches	85;	Conservative 54;	Mismatches 121;	Indels 144;
				Gaps 19;
Oy	43	VLRPTKDTLITGAPSSGAALPIRFIMLVPCAILFMI-----YAKLSNLSKQALFY	96	

```

Db      677   I S R D T K N ----- W K I I L C F I I P V G C G F V S F R K K P V D K H K L M Y       717
Qy      97    A V G - - - T P E L I - - - - - F F A L E P T V - I Y L B R D V L H P T E F A D R L Q A I L P G L I G V A     142
Dh      718   Y V A F P T S F P V F S M N V F Y I A F L L F A V Y L L M D H S V H P P E L V - - - - - L V S I V F       767
Qy      143   I L - - - - - R N W T F A A F Y I A E L M G S V M L S I - M E W G F A N E I T K I H E A K R F Y A L F G I G A N I S L     196
Dh      768   V L F C D E V A Q W Y V N G N V F P T D L M - N M D I L T G L Y F I A G I V F R L H - - - - - S S N K S S     815
Qy      197   L A S G R A I V M A S K L R A S V S E G V D P W C I S I R L M A M T I V S G L V M A S Y W I N K N V L T D P R F Y     256
Dh      816   L Y S G R V I E C - - - - - I D Y I I F T R L I H I F V - - - - -                               840
Qy      257   N P E M O K K G K A K P M N M K D S F L Y I D R S P Y I L L L V L A Y G I C I N L I - - - - -           304
Dh      841   - - - - - S R N I G P X I M L O R M L - I D V F F L F L P V M W V A F G A R O G I R O N E Q R W R I F     892
Qy      305   - E V T M K S J K L - - Q Y P N N D Y S E F M G N S F M W G V U S V L I M L V G G A V I K F - G M L I G A I V     360
Dh      892   R S V I E P L A M F G O V P S V D G T T Y P A C T F G N E S K P L C V A L D E H N L P R F E M I T I P L V     951
Qy      361   T P V W I L T G I V P A L V I F R N Q A S G I V A N E G T P L M L A V V G A I Q     404
Dh      952   C - I Y M L S T N I L V N L - - - - - L V A N F G Y I - - - - - V G T V Q       978

```

```

RESULT 23
US-09-328-352-4974
; Sequence 4974, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4974
; LENGTH: 419
; TYPE: PRF
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4974

```

Query Match	4.3%;	Score 112.5;	DB 4;	Length 419;
Best Local Similarity	20.0%;	Pred. No. 0.0017;		
Matches 94;	Conservative 61;	Mismatches 148;	Indels 167;	Gaps 22

```

OY      84  KLSNLSLQAL-----FAVGVPPILFFALPTVIVPLRDVLPHTEFADRLQALPP 135
Db      22  KLMTNGSTAFSLALAIAGFAIGI-----TFSS-----PM 52
OY      136  GLGLIVA-----IRNNVFAAFVYVLAIWGSVMTSLMFWGFANE----- 174
Db      53  GLPPIANDLGISIFPAGMLITGVALGVMLGAPFMTLWFGFGFARNALITLMAIFVGNL 112
OY      175  -----ITRIHEKAPFYALFGAGINISLASGRAIYMAKSLRASVE--- 215
Db      113  IAPSPNMSLLGALLITSUNG-----AFFGISGYA-----ASIVPAIKQASAVATMEM 163
OY      216  -----GVD--PW-----GISIRLIMAMTVISGLVIMASWYMINKNVLTDPREFYDEMO 262
Db      164  GLTIANIGVPLATVVGQINIGWRMSFLAISVIGIITMALM-----KALP 208
OY      263  KKKGAKPEKRMNNKDSFELYIDRSFYLLLTLLVIAIGICNLIIEVTWKSQKLQYPMNDY 322
Db      209  QGMWAQRK--NVKAEIKVLTFRPPVALLTLYVLGAGMFTLYTI-----APSLTEF 258
OY      323  SEPMKNSFMTGVSVLIMLFYGVGNVIRKFGMLTGALTVPVAVLLTGIVFFALVIFRNQA 382
Db      259  THASPTF-----ITFMVLVIGVFSFI-----GNHLGRFADISINKTLIGFIVL----- 302

```

QY 383 SGVAMECTPLMAVVG-AIONILSKTKALFDSTKEMAYTFLDQ-----QKVGK 436
DB 303 --LIVMTPTPLILQSOIGAAIALVIGCAATFALP-----PLQNRVNSVAHEADGL 352
QY 437 A-AIDVVARFGKSGGALIQOGLLV-----ICSGIAMPYLAIVL 477
DB 353 ASSVNIQAFNIGNAVGA--AAGALVLDLQKGSVSPFAGALLAGGLLV 400

RESULT 24

US-08-597-236-11
; Sequence 11, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesea
; APPLICANT: MOULET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; NUMBER OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-597-236-11

Query Match 4.2%; Score 111; DB 1; Length 360;
Best Local Similarity 17.2%; Pred. No. 0.002;
Matches 74; Conservative 80; Mismatches 142; Indels 134; Gaps 19;
QY 103 LIFPALFPTV--IYLRDVL-HPTFADRLQAILPPGLGLVAIRNWTFAFYLAELM 159
DB 2 VIFFLFPMIMITYLMTLLRQKQIQKTFICVLTFGLGFSASRASSVGTDTLYE-- 59
QY 160 GSVMLSMF-----WGFANEITKIHAKRFYALFGIGANISLASGRAIVMASKLRAS 212
DB 60 -NIFKSINYGISAENMGY-----VIYNKLIGSVFGY-----TGEHTTAANSVLIT 104
QY 213 VSEGVDPWGISLRLLMA-----MTIVSGIYLMASVWINKVLTDP 254
DB 105 ILIGFIWKVAEHVATFLYISLFYVATSFNISROFIAMGLVVAISFALDKKVM--- 160
QY 255 FYNPEEMQKKGAKPKNMKDSFLYLDKSPYILLTLTLVAYGICINLIEVTKSQKLT 314
DB 161 -----PWFILTVLTLPHATVAIVAFPYVLT--KV 188
QY 315 QYPMNDYSEF---MGNFSEFWTGVSVLIMLFVGANVIRKGMWLTG-----A 358
DB 189 HMDVKKTLISFPIITFAFIF-DALINIFVRFPPHYEM-----YITGQFNISDQGGGRV 242

QY 359 LVTEFVMTLNGIVPEFALVIFRNQASGLV-----AMFGTPLMLAVVGAIONILSKSTKY 413
DB 243 VLVKIFILL--LIFLFLFKSYALISBCHOSLALITVGLSIGVFNIL--LNRI 297
QY 414 ALPSTKEMAYTFLDQKVKGAIDVVARFGKSGA--LIQOGLVI-----C 462
DB 298 EMFYSILSIATFIP-----ALDYISLKFQKQAVRMLTIGLILLVLPVYIGVS 347
QY 463 GSGIAMPYVL 472
DB 348 GNVSGILPYV 357

RESULT 25

US-08-746-682A-11
; Sequence 11, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesea
; APPLICANT: MOULET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; NUMBER OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-746-682A-11

Query Match 4.2%; Score 111; DB 1; Length 360;
Best Local Similarity 17.2%; Pred. No. 0.002;
Matches 74; Conservative 80; Mismatches 142; Indels 134; Gaps 19;
QY 103 LIFPALFPTV--IYLRDVL-HPTFADRLQAILPPGLGLVAIRNWTFAFYLAELM 159
DB 2 VIFFLFPMIMITYLMTLLRQKQIQKTFICVLTFGLGFSASRASSVGTDTLYE-- 59
QY 160 GSVMLSMF-----WGFANEITKIHAKRFYALFGIGANISLASGRAIVMASKLRAS 212
DB 60 -NIFKSINYGISAENMGY-----VIYNKLIGSVFGY-----TGEHTTAANSVLIT 104
QY 213 VSEGVDPWGISLRLLMA-----MTIVSGIYLMASVWINKVLTDP 254
DB 105 ILIGFIWKVAEHVATFLYISLFYVATSFNISROFIAMGLVVAISFALDKKVM--- 160
QY 255 FYNPEEMQKKGAKPKNMKDSFLYLDKSPYILLTLTLVAYGICINLIEVTKSQKLT 314

QY 179 -----HEAKRFYALFGIANISL-LASGRAIVASKLRAVS-----EGVDP 219
 DB 137 LVIALAKRELMNDRAAAAAIIIGGATWALGIISSAAQLOQAKTSLPESLYNLGTVP 196
 QY 220 W--GISLRLLMAMTIVSGVLMA-SYWMINK--NVLT----DPRFNPPEMOKGKKAKP 270
 DB 197 FETETFLMOSIMTIIIVLSIAIVWASAPKNSVKTIDSPVOQF--BEETKNMAKSTRP 254
 QY 271 KKNMDSFLYDRSPYLLITLLVAVGICINLIEVTWKSQKLOQYPMNDYSEFMGNS 330
 DB 255 GD-----WLENSP---LITIIIVVLGIMMFEERS-KSNPIIAISSINTY-----NEV 298
 QY 331 F 331
 DB 299 F 299

RESULT 30

US-09-328-352-6605
 ; Sequence 6605 Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6605
 ; LENGTH: 548
 ; TYPE: PRP
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-6605

Query Match 4.1%; Score 108.5; DB 4; Length 548;
 Best Local Similarity 22.0%; Pred. No. 0.007;

Matches 105; Conservative 58; Mismatches 140; Indels 175; Gaps 27;

QY 99 GTFFLIFALFPTVYPLRDVLP--TEPADRLQAILPGLGLVAILRNWTFAPFYVLA 157
 DB 80 GTTNVAFSILFAIILITLGI--PLAVYARVN-----IDLDIRGAGG--YI--- 125
 QY 158 LMGSVMLSMFEGFANEITKTHEAKRFYALFGIANISLILASGRAI---VMASKLRAS 212
 DB 126 --GSVLTSLIFASF--TFI-----FPALIG--SMAQGLLGLGIPLMAGVLIST 169
 QY 213 VSEGVDP--WGI-SLRLLMAMTIVSGVLMA--ASYWMINKNVLTDPFNPPEMOKGK 265
 DB 170 VM--VILVLYGMKALSKLOVWTTPLMLVLMIGVAVL-----IYOEPLVSOFAFTTGH 222
 QY 266 KGAKRPMNMKDSFLYDRSPYLLITLLVAVGICINLIEVTWKSQKLOQYPMNDYSBF 325
 DB 223 EGFAP-----VDMAIMLGAGICISL-----MQIGEQDIYARF 256
 QY 326 M-----GNFSFWTGVSV-----LIMFVGSNVTRKF----- 352
 DB 257 MPAKTKENSKAWMAAVISAGPGVILGAIKQIIGAFGLFYLTJKIPGVNSTEPVOQFNA 316
 QY 353 -----GMLTGALVTPVAVL-----LTGIYFPALVIFR 379
 DB 317 FHDMLPGMAALTLVAVLIVISQIKINVTNAYSGLAMTSAYTRISKHPRGVIVFIV--- 373
 QY 380 NQASGLVAMFGTTPMLAVVVG-----AIQNILSKST-----KVALPSTKEMAYIPLDQ 429
 DB 374 NLAIALALMBEGMFAVLGKIIGFYSNFAIAMVVAVATDISINKYVLKLSKPEPR-RDM 432
 QY 430 EOKVK--GKAIDV-----VAARFKSGGALIQGGLVYICSGIGANTPYLAVILFITI 480
 DB 433 LYNVAVGVAVFLVSGAGLSIAAFGLIGSFL-----ADYSPILAVLAVFL 478

RESULT 31

US-09-555-313B-2
 ; Sequence 2, Application US/09555313B
 ; Patent No. 6506580
 ; GENERAL INFORMATION:
 ; APPLICANT: PICHMEISTER, Rudolph et al.
 ; TITLE OF INVENTION: Splicing variants of the human serotonergic receptor
 ; TITLE OF INVENTION: 5-HT₄ and uses thereof, in particular for screening
 ; FILE REFERENCE: P06762US00/BAS
 ; CURRENT APPLICATION NUMBER: US/09/555,313B
 ; CURRENT FILING DATE: 2002-08-13
 ; PRIOR APPLICATION NUMBER: FR 97/15037
 ; PRIOR FILING DATE: 1997-11-28
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 760
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-555-313B-2

Query Match 4.1%; Score 108.5; DB 4; Length 760;
 Best Local Similarity 21.8%; Pred. No. 0.011;

Matches 62; Conservative 54; Mismatches 93; Indels 75; Gaps 15;

QY 269 KPKMMKDSFLYDRS-----FYILLITLL--VIANGLINL--IEVTWKSQKLOQYPM 319
 DB 378 KPSMDKLDANVSSEEGFSGVSEKVVLTFLSTVILMAIIGNLVAVAVCMRQRLRIKTY 437
 QY 320 NDYSEFMGNFSPFWTGVSVLIMLFGVGNVIRKFGMLTGALV---TPVMVLLTGIVFPAL 375
 DB 438 -----FIVSLAPADLDVSVLWMPFGAIELVQDI-WYGEVVCVTRTSIDVLLTASITHL 491
 QY 376 -----VIFRQASGLVAMFGTTPMLAVVVG--AIQNILSKSTYALFD 417
 DB 492 CCISLDRYAICCOPLVYRNK-----MTPLRLALMLGCGWVIFPFIS----- 533
 QY 418 STKEAAYIPLDQEQVKVKKALIDVYAR-FKSGGA-----LIQGLVIVGSGIGAMTPY 471
 DB 534 -----FLPIMQGMNID--IIDLIEKRFQNSNSTYCVFMPKRVATCSVAFYIFP 585
 QY 472 LAVILLFIAIMVLSATLKLFLAQSALKEQEVNQAQDSAPASS 515
 DB 586 LLMVLAV-YRIYVTRAKHAHQIQMLQRA-----GASSSESRFQSA 623

RESULT 32

US-09-252-991A-24105
 ; Sequence 24105 Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: ABRUSIONA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 24105
 ; LENGTH: 816
 ; TYPE: PRP
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24105

Query Match 4.1%; Score 108.5; DB 4; Length 816;
 Best Local Similarity 19.3%; Pred. No. 0.013;
 Matches 57; Conservative 56; Mismatches 99; Indels 83; Gaps 10;

QY 134 PPGLLGVAILRNWTFAPFYVLAELMGSVMLSMFEGFANEITKTHEAKRFYALFGIAN 193

```
Db      | 410 PGAGLIFP-----VASVIAAILALFPGY-----LS 438
Qy      | 194 ISLASGRAIIVASKLRASVSEGVDPWGISLRILMAMTVISGLVIMASYMINKVLTDP 253
Db      | 439 LAVFLTKR-LMWISVIAATTVILLAFPGDICEFLISPKRPSGLAL-----ANAL--- 486
Qy      | 254 RRYNEBEMOKGKKGAPEKPKAMKDSFLYDRSPYIILLTLVLVAYGICINLIEVTKSQJK 313
Db      | 487 -----GSPRHQAQASTYLAGVGTLLTLTLTLAF-----LPVGSSEELL 528
Qy      | 314 LQYPMNDYSEFMGNFSFWTGVSVLIMFVGG-----NYIRKFGMLTGALVTPV----- 363
Db      | 529 QGFAGLGESSRLGNINIVPQDILVALLLFVGGLFRLRYKR--WLSERLLBETNMDACM 586
Qy      | 364 ---MVLITGIVFPALVIFRNQASGLVAMFGTTPMLAVVG-----AIONILS 408
Db      | 587 RASLVTLVGYLGVLLAMLMWSTLRINTSLFWVVSALSVGIGFGLQAIQHFIS 641

RESULT 33
US-09-252-991A-29935
; Sequence 29935, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29935
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29935

Query Match      4.1%; Score 108; DB 4; Length 463;
Best Local Similarity 18.9%; Pred. No. 0.0061;
Matches 79; Conservative 66; Mismatches 171; Indels 102; Gaps 14;

Qy      | 139 GLVALIRNMTFAAFYVLAELMGSVMLMFNGFANEITKIHAKRYALFGIGA-----N 193
Db      | 96 GRVRYV-TWTIVLFSIGTALIG--LTQNYWQFA-----LIRVASLIGLSLFAVACN 143
Qy      | 194 I-----SLASGRAIIVASKLRASVSEGVDPWGISLRILMAMTVISGLVIMAS 241
Db      | 144 IIMSECVGTFRYRTTILAAQAGMTVGYLVATLLAGQIIFEGWRALFTAVAPALVCLAL 203
Qy      | 242 YVMINKNVLTDRFYNPEE-----MOKGK--GAKPKKM-----MKDS 277
Db      | 204 RFWV-----PESPFWLAAQAGKORPEAQAQSHAEAGAGPLGRMLGDEPTRGM 251
Qy      | 278 FLYLDRSPYIILLTLVLVAYGICINLIEVTKSQJLQYPMNDYSEFMGNFSFWTGVSV 337
Db      | 252 FLL-----WSLNSCPLLPGYGTNNMMPYSLESLGNMKNSTGY-----WGTYS 297
Qy      | 338 VLIIMLFVGNVIRKFGMLTGALVTPVWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLA 397
Db      | 298 AMILKQVAGVACAD---LFGRLTFEAGAGISAAFWPLLVFVHSPENIIMILITGFIYG 354
Qy      | 398 VVVVGAIONLSTKSTKALFDSTKEMAYIILDOEQYKKGKALDVVAARSGSGALLIOG 457
Db      | 355 VVGVVATVYTESF-----STRYRGSV--GTAYNIGRIGAIVAPRG 394

Qy      | 458 ILVIGSGISGAMPYLAIVLLFIATLWVSATKLNKFLAQSALKEQEVAAQESAPASS 515
Db      | 395 IGLLATHVSIAGAGFLVMGITYVLTGLIPALFTADRLIYDNRERSRQAESRQAQESA 452
```

```
RESULT 34
US-09-328-352-5089
; Sequence 5089, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5089
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5089

Query Match      4.1%; Score 107.5; DB 4; Length 499;
Best Local Similarity 21.9%; Pred. No. 0.0077;
Matches 95; Conservative 57; Mismatches 151; Indels 131; Gaps 23;

Qy      | 34 FPCIFENVTVLRDFTDLIVGAPSGAIPAIFIKFVLVPC-AIFMLIY-----AKLS 86
Db      | 135 FF--FN-TYIQTTR-----ATGFDLPVGH-LFWPIYGIWALTLIILYGVKKGVLS 183
Qy      | 87 NILSKQALFYAVGTPFLIFALFPTVIYPLRDVILHPTFADRLQAILPPGLGLVAIIR- 145
Db      | 184 NKIFWPLIF-----ILFTIL--VIQSLR-----LPGAVQGLNAFFTP 218
Qy      | 146 NWTFAFY-VLAELMGSMLSL-----MNGFANETKTHE 180
Db      | 219 NWSAMMDYKWLAAAGHTFFSLSVGFMVTYASYLAKRKTNLTGSLGLVGFANASTELIA 278
Qy      | 181 AKRFYALFGI-----GANISLAS--GRATVASKLRASVSEGVDPWGISLRILMAMTI 232
Db      | 279 GIGIFALGFMAHAAGTIVKDVSGIGLAFIAPFKIISLSGADLFGI---LFFSLF 335
Qy      | 233 VSGLVIMASYMINKNVLTDRFYNPEEMOKGKKGAKPKMNMKSF--LYLDRSPYIILL 290
Db      | 336 VAGISMSVS-----LLEVPLAMQDKLKWGRKKAVTIIGGSAIVSIIIFSSVNAIKL 388
Qy      | 291 TLUV-----IAYGICINLIEVTK--KSQLKQYPMNDYSEF--MGNFSFWTGVSVL 339
Db      | 389 VDIYDHTFNINIGIIGALLSTISVAMFRRSALKELRDVNRISTIQDK----- 437
Qy      | 340 IMLFVGNVIRKFGMLTGALVTPVWVLLTGIVFPALVIFRN-----QASGLVAMFGTTPLM 395
Db      | 438 -----GMDFTLTYITSLIILITLCMTVFNLIKNGYDYSLSGQVFGMSVI 484
Qy      | 396 LAVVGAIONILSK 409
Db      | 485 FCAVVAI---VLSK 495

RESULT 35
US-09-328-352-4168
; Sequence 4168, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4168
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
```


Db 20 VLLTFLSTVILMALIGLWVAVCMWRQLKIKTNY-----FIVSLAFADLVSVLVM 74
QY 342 LFVGNVIRKRGMLTGLV-----TPVMVLLTGIVFPAL-----VIFRQA 382
Db 75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLITIASIFHLCCISLDRYVAICCPVLYRNK- 132
QY 383 SGLVAMFGTPTPLMAVVG---AIONILSKSTKVALFDTSEKMAVYIPLDQEQVKGKAI 439
Db 133 -----MTPLRIALMGCGCWVITPTIS-----FLPIMQGNNIG--II 167
QY 440 DVNAAR-FGKSGA-----LIOGGLVYICSGIGAMTPYLAIVLLFIITAIWLSATKML 493
Db 168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
QY 494 FLAOSALKEQVAVQEDSAPASS 515
Db 227 QMLORA-----GASSESRPOSA 243

RESULT 38
US-09-555-313B-24
Sequence 24, Application US/09555313B
Patent No. 6506580
GENERAL INFORMATION:
APPLICANT: FISCHMEISTER, Rudolph et al.
TITLE OF INVENTION: Splicing variants of the human serotonergic receptor
FILE REFERENCE: P067620500/BAS
CURRENT APPLICATION NUMBER: US/09/555,313B
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: FR 97/15037
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 380
TYPE: PRT
ORGANISM: Homo sapiens
US-09-555-313B-24

Query Match 4.0%; Score 105.5; DB 4; Length 380;
Best Local Similarity 22.1%; Pred. No. 0.0085;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

QY 287 ILLTLL--VIAVIGICNL---IEVTWKSQKLQYPNNDYSEFGNFSFMTGVSVLIM 341
Db 20 VLLTFLSTVILMALIGLWVAVCMWRQLKIKTNY-----FIVSLAFADLVSVLVM 74
QY 342 LFVGNVIRKRGMLTGLV-----TPVMVLLTGIVFPAL-----VIFRQA 382
Db 75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLITIASIFHLCCISLDRYVAICCPVLYRNK- 132
QY 383 SGLVAMFGTPTPLMAVVG---AIONILSKSTKVALFDTSEKMAVYIPLDQEQVKGKAI 439
Db 133 -----MTPLRIALMGCGCWVITPTIS-----FLPIMQGNNIG--II 167
QY 440 DVNAAR-FGKSGA-----LIOGGLVYICSGIGAMTPYLAIVLLFIITAIWLSATKML 493
Db 168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
QY 494 FLAOSALKEQVAVQEDSAPASS 515
Db 227 QMLORA-----GASSESRPOSA 243

RESULT 39
US-08-446-822-8
Sequence 8, Application US/08446822
Patent No. 5766879
GENERAL INFORMATION:
APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
TITLE OF INVENTION: DNA ENCODING 5-HT4 SEROTONIN RECEPTORS
TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,822
FILING DATE: June 1, 1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: White, P., John
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPM/MAT
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-822-8

Query Match 4.0%; Score 105.5; DB 1; Length 388;
Best Local Similarity 22.1%; Pred. No. 0.0087;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

QY 287 ILLTLL--VIAVIGICNL---IEVTWKSQKLQYPNNDYSEFGNFSFMTGVSVLIM 341
Db 20 VLLTFLSTVILMALIGLWVAVCMWRQLKIKTNY-----FIVSLAFADLVSVLVM 74
QY 342 LFVGNVIRKRGMLTGLV-----TPVMVLLTGIVFPAL-----VIFRQA 382
Db 75 PFGAIEIVQDI-WIYGEVFCIVRTSLDVLITIASIFHLCCISLDRYVAICCPVLYRNK- 132
QY 383 SGLVAMFGTPTPLMAVVG---AIONILSKSTKVALFDTSEKMAVYIPLDQEQVKGKAI 439
Db 133 -----MTPLRIALMGCGCWVITPTIS-----FLPIMQGNNIG--II 167
QY 440 DVNAAR-FGKSGA-----LIOGGLVYICSGIGAMTPYLAIVLLFIITAIWLSATKML 493
Db 168 DLIERKRNQNSNSTYCVFMVNRKPYAITCSVVAFYIPFLMWLAY-YRIYVTAKEHAHQI 226
QY 494 FLAOSALKEQVAVQEDSAPASS 515
Db 227 QMLORA-----GASSESRPOSA 243

RESULT 40
US-09-328-314-8
Sequence 8, Application US/09328314
Patent No. 6331401
GENERAL INFORMATION:
APPLICANT: Gerald, Christophe
APPLICANT: Hartig, Paul R.
APPLICANT: Branchek, Theresa
TITLE OF INVENTION: DNA Encoding 5-HT4 Serotonin Receptors And Uses Thereof
FILE REFERENCE: 42667-A2-PCT-US
CURRENT APPLICATION NUMBER: US/09/328,314
FILING DATE: 1998-04-03
EARLIER APPLICATION NUMBER: 08/446,822
EARLIER FILING DATE: 1995-07-31
EARLIER APPLICATION NUMBER: PCT/US93/12586
EARLIER FILING DATE: 1993-12-22

EARLIER APPLICATION NUMBER: 07/996,772
EARLIER FILING DATE: 1992-12-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 388
TYPE: PRF
ORGANISM: Homo sapiens
US-09-328-314-8

Query Match 4.0%; Score 105.5; DB 4; Length 388;
Best Local Similarity 22.1%; Pred. No. 0.0087;
Matches 58; Conservative 50; Mismatches 83; Indels 71; Gaps 14;

QY 287 ILLTL--VIAVCINL--IEVTKSOLKQYNNNDYSEFMGNSEFWTGVSVLIM 341
DB 20 VVLLFPLSTVIMAILGNLLMVAVCGMDROLRIKNTY----FVSLAFADLVSVLVM 74
QY 342 LFGGVNIRKFGMLTGALV---TPVMVLLTGIVFPAL-----VIFRNOA 382
DB 75 PRGALBLVODI-WYGEVFCVLTSLDVLTTASTFHLCCISLDRYAICQPLVYRNK- 132
QY 383 SGLVAMFGTTPMLAVVG--AIONILSKSTKYLFDSTKEMAYIPLDQOKVKGKAI 439
DB 133 -----MTPLRIMLMGGCWVLPFEIS-----FLPIWQGNNG--II 167
QY 440 DVVAAR-FKSGGA-----LIOGLLVTCISGANTPYLAVILLPTIAMLVSAIKNL 493
DB 168 DLIERKRNQNSNSTYCVFMVKNRPYATCSVVAFYIPLMLVLAAY-YRIYVTAKENHAI 226
QY 494 FLAQSALKQEVAVOEDSAPASS 515
DB 227 QMLQRA-----GASSESRQSA 243

Search completed: November 25, 2003, 10:17:11
Job time : 26 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:16:18 ; Search time 54 Seconds

(without alignments)
1759.042 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630

Sequence: 1 MTKTEKPKFGKLSFLWPIH.....AQSLAKGEVAQEDSAPASS 515

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 18443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 3: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 18: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2630	100.0	515	9 US-09-892-851-2	Sequence 2, Appl1
2	162	6.2	538	10 US-09-738-626-3617	Sequence 13617, Ap
3	132.5	5.0	463	9 US-09-815-242-10144	Sequence 10144, A
4	129	4.9	462	9 US-09-815-242-10161	Sequence 10161, A
5	125	4.8	587	14 US-10-173-519-5	Sequence 5, Appl1
6	125	4.8	587	15 US-10-167-994-12	Sequence 12, Appl1
7	120.5	4.7	713	14 US-10-014-717-11	Sequence 11892, A
8	120.5	4.6	395	9 US-09-815-242-13892	Sequence 13892, A
9	120.5	4.6	425	9 US-09-815-242-12171	Sequence 12171, A
10	118	4.5	466	9 US-09-815-242-13831	Sequence 13831, Ap
11	117.5	4.5	430	9 US-09-815-242-5678	Sequence 5678, Ap
12	116.5	4.4	477	9 US-09-815-242-13840	Sequence 13840, A
13	115.5	4.4	410	10 US-09-738-626-3861	Sequence 3861, Ap
14	115	4.4	392	12 US-10-181-319-4	Sequence 4, Appl1
15	115	4.4	708	10 US-09-738-626-5024	Sequence 5024, Ap

16	113	4.3	1095	9 US-09-759-143-778	Sequence 778, App
17	113	4.3	1095	9 US-09-780-669-778	Sequence 778, App
18	113	4.3	1095	9 US-09-822-827-778	Sequence 778, App
19	113	4.3	1095	10 US-09-895-793-778	Sequence 778, App
20	113	4.3	1095	10 US-09-895-814-778	Sequence 778, App
21	113	4.3	1095	12 US-10-144-678A-778	Sequence 778, App
22	113	4.3	1095	12 US-10-294-025-778	Sequence 778, App
23	113	4.3	1095	14 US-10-012-896-778	Sequence 778, App
24	113	4.3	1095	15 US-10-205-823-421	Sequence 421, App
25	113	4.3	1268	10 US-10-171-319-11	Sequence 11, Appl1
26	113	4.3	1681	10 US-09-920-653-3	Sequence 3, Appl1
27	111.5	4.2	1083	10 US-09-738-626-6914	Sequence 6914, Ap
28	111	4.2	741	10 US-09-738-626-4455	Sequence 4455, Ap
29	111	4.2	1095	9 US-09-759-143-780	Sequence 780, App
30	111	4.2	1095	9 US-09-780-669-780	Sequence 780, App
31	111	4.2	1095	9 US-09-822-827-780	Sequence 780, App
32	111	4.2	1095	10 US-09-895-793-780	Sequence 780, App
33	111	4.2	1095	10 US-09-895-814-780	Sequence 780, App
34	111	4.2	1095	12 US-10-144-678A-780	Sequence 780, App
35	111	4.2	1095	12 US-10-294-025-780	Sequence 780, App
36	111	4.2	1095	14 US-10-012-896-780	Sequence 780, App
37	110.5	4.2	425	9 US-09-815-242-10467	Sequence 10467, A
38	110.5	4.2	425	12 US-10-287-274-428	Sequence 428, App
39	110.5	4.2	476	10 US-09-738-626-5741	Sequence 5741, App
40	110	4.2	927	12 US-10-032-585-7923	Sequence 7923, App
41	109.5	4.2	665	10 US-09-942-447-2	Sequence 2, Appl1
42	108.5	4.1	511	15 US-10-156-761-11421	Sequence 11421, A
43	108.5	4.1	587	9 US-09-815-242-5694	Sequence 5694, App
44	108.5	4.1	596	9 US-09-815-242-5244	Sequence 5244, App
45	108.5	4.1	603	9 US-09-815-242-12667	Sequence 12667, A
46	108.5	4.1	620	9 US-09-815-242-12525	Sequence 12525, A
47	108	4.1	320	15 US-10-156-761-9297	Sequence 9297, App
48	108	4.1	451	9 US-09-815-242-5079	Sequence 5079, App
49	108	4.1	933	9 US-09-815-242-11817	Sequence 11817, A
50	107	4.1	527	10 US-09-881-752A-136	Sequence 136, App
51	107	4.1	1043	10 US-09-738-626-6647	Sequence 6647, App
52	106.5	4.0	475	10 US-09-738-626-6236	Sequence 6236, App
53	106	4.0	530	15 US-10-156-761-13471	Sequence 13471, A
54	105.5	4.0	360	12 US-09-826-509-443	Sequence 443, App
55	105.5	4.0	378	12 US-09-826-509-445	Sequence 445, App
56	105.5	4.0	380	12 US-09-826-509-441	Sequence 441, App
57	105.5	4.0	387	12 US-10-118-661-23	Sequence 23, Appl1
58	105.5	4.0	387	12 US-09-826-509-436	Sequence 436, App
59	105.5	4.0	388	9 US-09-989-861-8	Sequence 8, Appl1
60	105.5	4.0	388	12 US-09-826-509-439	Sequence 439, App
61	105.5	4.0	388	15 US-10-157-031-123	Sequence 123, App
62	105.5	4.0	388	15 US-10-157-031-124	Sequence 124, App
63	105.5	4.0	388	15 US-10-225-567A-18	Sequence 18, Appl1
64	105.5	4.0	489	15 US-10-156-761-11766	Sequence 11766, A
65	105.5	4.0	502	15 US-10-156-761-8917	Sequence 8917, App
66	105.5	4.0	531	12 US-10-154-419-81	Sequence 81, Appl1
67	105.5	4.0	531	12 US-10-146-733-31	Sequence 31, Appl1
68	105	4.0	427	10 US-09-922-501-6	Sequence 6, Appl1
69	105	4.0	451	10 US-09-738-626-3815	Sequence 3815, App
70	105	4.0	490	9 US-09-739-457-3	Sequence 3, Appl1
71	105	4.0	490	9 US-09-739-457-6	Sequence 6, Appl1
72	105	4.0	490	9 US-09-739-457-7	Sequence 7, Appl1
73	105	4.0	490	12 US-09-849-138-28	Sequence 28, Appl1
74	105	4.0	490	12 US-09-849-138-43	Sequence 43, Appl1
75	105	4.0	643	16 US-10-080-170-626	Sequence 626, App
76	105	4.0	1104	11 US-10-171-319-8	Sequence 8, Appl1
77	105	4.0	2261	11 US-09-984-827-132	Sequence 132, App
78	104.5	4.0	786	9 US-09-815-242-10103	Sequence 10103, A
79	104	4.0	479	12 US-10-338-075-937	Sequence 937, App
80	103.5	3.9	572	10 US-09-919-781-2	Sequence 2, Appl1
81	103.5	3.9	572	12 US-10-154-419-13	Sequence 13, Appl1
82	103.5	3.9	772	10 US-09-738-626-6661	Sequence 6661, App
83	103	3.9	410	9 US-09-767-041-20	Sequence 20, Appl1
84	103	3.9	438	9 US-09-815-242-5203	Sequence 5203, App
85	103	3.9	469	9 US-09-925-302-627	Sequence 627, App
86	103	3.9	537	15 US-10-156-761-12236	Sequence 12236, A
87	103	3.9	550	15 US-10-272-613-8	Sequence 8, Appl1
88	103	3.9	2261	11 US-09-984-827-128	Sequence 128, App


```
Qy 332 WTGVSVLIMFVGGNVIRKFGMLTGLVTPVMTLLTGIVFPAVIFRNQASG--VAMF 389
Db 370 -KDMIGFLVLAIFILQGFALFNM-----TIGTWAV--QGAIGLEAIGIT 412
Qy 390 GTPEMLAVVGAION--ILSKSTKYALFDSTKEMAYIPLDQEKVKKAAIDVVAARFG 447
Db 413 GFPAIIAIFIIILASCLNLIISGSAMWTL-----MAAVFVPMFALLGYEPEFIOAARVFG 466
Qy 448 KSGGALL---OQGLVYTGSGIGAMTP-----YLAIVLLFIATWVSATKLNKLFIAQS 498
Db 467 DSAQTQVITPLNPMYIVILGLRRYEPDAGLTLMKRLIPVIPEMLAWATLAIWIFYADL 526
Qy 499 AL 500
Db 527 PL 528

RESULT 3
US-09-815-242-10144
; Sequence 10144, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10144
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10144

Query Match 5.0%; Score 132.5; DB 9; Length 463;
Best Local Similarity 22.3%; Pred. No. 0.00058;
Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;

Qy 69 WLVPVCAIIFMLIYAKLSNI--LSKQALFYAVGTPELIFALFPYIYPLRDVLPHTERFAD 127
Db 39 WISVPCILLAFCAVWMLSAVAVNLPKVGFNFTTQDLFMLTALPSVSGALLRV--PYSF-- 94
Qy 128 RLQALIPGLGLGVAIIL--RNMTPAFVYVLAELMGSVLGLMPFGFAIEIRKIHAKRFY 185
Db 95 -----MPIFGGRMT--ARST-----GILIIPCVWIGFA-----VQDTSTPY 130
Qy 186 ALFPIGAINISLLASGRAIYVASKLRASVS-----EGVDPMGISLRLLMAM 230
Db 131 SVFTI---ISLCCGFAGANFASSM-ANISFFPKQKQCGALGLNGCIGNMGVSVWQIVAP 186
```

```
Qy 231 TIIVSGLVIMASVWMIINKVLTDPFRFYNPEEMQKKGAKPRGMKKDSFLYDRSPYILL- 289
Db 187 LVVS-LSTFAVF-----DSQVKQPDGIE--LYIANASWIVWP 221
Qy 290 -LTLVIAVAGICINLIEVYTKSQKLQYNNND-----YSEFMGNF-SFWTGV--- 335
Db 222 FLAFTIAVAGCNDL-ATSKASIKEDLPYKRGHLMIMSLYLATSGSFIGSAGFAML 280
Qy 336 -----VSULIMFVG--GNVIRKFGW-----LTGALVTPV---MWLTGIVFPLV 376
Db 281 SKTQFPDVQILQVAFPGFPGIATARSAGALSDRLGTRVTLVNFILMAIFSGILFLTLP 340
Qy 377 IFRNQASGLVAMFGTPEMLAVV-----GAIGNILSKSTKYALFDSTKEMAYIPLDQE 430
Db 341 T-DGQGSFMAFPA--VFLLALFLTAGLSGSTFQMTSVIFRKLTMDEKAEG---GSD 392
Qy 431 QKVKAIDVVA-----RFGKSGALLQGL---LVICGS-IGAMTPYLAIVLLFIIA 481
Db 393 ERARREATDPAALGFIISALGATGFFIPKAFQSSIALTGSFVGAMKVFILFIYACVI 452
Qy 482 IWLV 485
Db 453 TWAV 456

RESULT 4
US-09-815-242-10161
; Sequence 10161, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10161
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10161

Query Match 4.9%; Score 129; DB 9; Length 462;
Best Local Similarity 22.3%; Pred. No. 0.0012;
Matches 108; Conservative 71; Mismatches 171; Indels 134; Gaps 26;

Qy 69 WLVPVCAIIFMLIYAKLSNI--LSKQALFYAVGTPELIFALFPYV-----IY 114
Db 37 WISVPCILLAFCAVWMLSAVAVNLPKVGFNFTTQDLFL--TLAPSVSGALLRVYSPMV 94
```

```

QY 115 PLRDLVHPTREPADRLQALIPGLGLVALIRNMTFAAFYVLAELMGVSLMFNG--FA 172
DB 95 PLFGRRMTVFSTAL-LIIPCVMGLIAVONPPTFGIFVLALGCG-----FAGANPA 146
QY 173 NEITKIH-----BAKRFYALFGIGANISLASGRAIWMASKLRASVSEGVDPWGISLRLM 228
DB 147 SSMGMSISFFPKAKGSGAL-GINGGL-----GNLGVSMQVAPLIVIFVPPFA 193
QY 229 AMTI-----VSGVLVMAASYW-----WINKNVLTDRPFYNPEEMQKKGA 268
DB 194 FLGVGVGPOADGSVMSLANPAAMIWVPLAATIAMSGNNDI-----A 236
QY 269 KPXNMKMSFLYLDSPYILLTLVIA-YGICILIEVTWKSQKLQYPMNN-----DY 322
DB 237 SSRASIAOQLPYLOPL-HLMILSLLYLATFGSGFIG-PSAGFAMLAKTOFPDVNILLRLAF 294
QY 323 SEFMNFSFWTGVSVLIMLFVGNVIRKFGMLTGALVTPVWVLL-----TG-- 369
DB 225 GPFIG-----ALARSVGGAISDKFGGVRVTLINFIEMALFSALEPILPCTGSG 343
QY 370 --IYFALVIFENQASGLVAMFGTTPMLAVVGAIONILSKSKY--ALFDSKEMAYI 425
DB 344 NPIAFYAVFMGLFLTAGLGS--GSTFOMIAVIFROI-----TIYRVMMKGSDEQAH- 393
QY 426 PLDOQKTKGKRAIDIVNARFKSGGALIQGL--LVICGS-IGAMTPYLAIVILLFIIA 481
DB 394 --KAAVTETAALGFLISA-IGAVGGFFLPOAFGMSLNMWTSPPVAMKVFILFYIVCVLL 449
QY 482 IWLIV 485
DB 450 TWLV 453

```

```

RESULT 5
US-10-173-519-5
; Sequence 5, Application US/10173519
; Publication No. US20020193582A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 69624, A Human Transporter Family Member
; TITLE OF INVENTION: and Uses Therefor
; FILE REFERENCE: MP101-098P1RKM
; CURRENT APPLICATION NUMBER: US/10/173, 519
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/298, 970
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 587
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-173-519-5

```

```

Query Match 4.8%; Score 125; DB 14; Length 587;
Best Local Similarity 21.2%; Pred. No. 0.0041;
Matches 116; Conservative 67; Mismatches 179; Indels 184; Gaps 23;
QY 67 KFWLVVPCAIIFML-----IYAKLSNLSKQALFYAVGTFFLFFPALFFPVYIPL 116
DB 12 RFLYIVLCLPFLPLPLIVQTKKAYCAVSIIL--MALIMCTEALPLAVTALFFIVLFFPL 69
QY 117 RDVLAHPT-----PADRLQALIPGLGLVALIRNMTFAAFYVLAELMGVSLMFNGF 171
DB 70 MGINDASVECEYKFD--TNILFVGGLMVAIAV-----EHM----- 103
QY 172 ANEITKIHAKRFYALFGIGANISLASGRAIWMASKLRASVSEGVDPWGISLRLMAMT 231
DB 104 -----NLHKRIALQVLLIIGVRPALLLLGFWLVTAFA-----LSMW--ISNTATTAMM 148
QY 232 IVSGVLVMAASYWINKV---LTDPRFYNPE-----MQKKK 266

```

```

DB 149 VPIGHAVLEQVQSGKQVGGNNPTEFLQSECCQKVEYTKLDNGQPVASAPSEPRPTQKQE 208
QY 267 GAKPRNMKMSFLYLDSPYILLTLVIAVGICINLIEVTWKSQKLQYPMNDYSFEM 326
DB 209 HHRFSQGLSLCICYSASIGGIATLT-----GTPPNVLQGVNSLPQNGNVNFAWGF 262
QY 327 GNSEFMGVSVL-----MLFVGNVIRKFGM-----LTGAL- 359
DB 263 GFAFPTWILLIAMMLQVLFVGNFRKMFGEGBEGBEKKOAFQVYKTYQLLGPMS 321
QY 360 -----VTPVWVLLTGIVF-----FALVIFENQASGLVAMFGTTPMLAVVGAION 405
DB 322 FAEKTVTLFVLVWVMTREPRGFPFGGDTVFANEGQSWASDGTVAIFSLVWFILP- 380
QY 406 ILSKSTKALFDSKEMAYIP-LDOQKVKGA-----AIDVNAAR 445
DB 381 -----SKIFGLMQDPKPKGLKAPPAILTWTVDMPNNIVILLGG 423
QY 446 FKGSGALIQGLLVICSGISGAMTPY-----LAVILLEFIATML-----VSATKLNKL 493
DB 424 FALKGS--EBSGLSEWIGD--KLTPLOHIPSATRAVILCLIALFTECTSNVATTYLLP 480
QY 494 FLAQSA 499
DB 481 ILASMA 486

```

```

RESULT 6
US-10-167-994-12
; Sequence 12, Application US/10167994
; Publication No. US20030082647A1
; GENERAL INFORMATION:
; APPLICANT: Reenan, Robert A.
; APPLICANT: Rogina, Blank
; APPLICANT: Helfand, Stephen L.
; TITLE OF INVENTION: TRANSPORTER PROTEIN
; FILE REFERENCE: 13407-013001
; CURRENT APPLICATION NUMBER: US/10/167, 994
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/255, 013
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 10/017, 479
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 587
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-167-994-12

```

```

Query Match 4.8%; Score 125; DB 15; Length 587;
Best Local Similarity 21.2%; Pred. No. 0.0041;
Matches 116; Conservative 67; Mismatches 179; Indels 184; Gaps 23;
QY 67 KFWLVVPCAIIFML-----IYAKLSNLSKQALFYAVGTFFLFFPALFFPVYIPL 116
DB 12 RFLYIVLCLPFLPLPLIVQTKKAYCAVSIIL--MALIMCTEALPLAVTALFFIVLFFPL 69
QY 117 RDVLAHPT-----PADRLQALIPGLGLVALIRNMTFAAFYVLAELMGVSLMFNGF 171
DB 70 MGINDASVECEYKFD--TNILFVGGLMVAIAV-----EHM----- 103
QY 172 ANEITKIHAKRFYALFGIGANISLASGRAIWMASKLRASVSEGVDPWGISLRLMAMT 231
DB 104 -----NLHKRIALQVLLIIGVRPALLLLGFWLVTAFA-----LSMW--ISNTATTAMM 148
QY 232 IVSGVLVMAASYWINKV---LTDPRFYNPE-----MQKKK 266
DB 149 VPIGHAVLEQVQSGKQVGGNNPTEFLQSECCQKVEYTKLDNGQPVASAPSEPRPTQKQE 208
QY 267 GAKPRNMKMSFLYLDSPYILLTLVIAVGICINLIEVTWKSQKLQYPMNDYSFEM 326

```

```

Db      209 HHRFQGLSLCTCYASISGIGIATLT-----GTTFNLVLQGGVNSLFPQNGVNVFASMF 262
Qy      327 GNFSGWTVGVSVLT-----MLFVGVNVIKRGW-----LTGAL- 359
Db      263 G-FAFPTMIILLLLAMLMQVLFGLVFNFRKNFGFGEGERKQAAFOVIKQYRLIGMS 321
Qy      360 -----VFVWVLLTGTIVF-----FALVIFRQASGLVAMFGTDTMLAVVGAION 405
Db      322 FAEKTVTLFVLVLVMTTRREGFPFGMDIVFANEKQSMASDGTVAIFISLVMTIP- 380
Qy      406 ILKSTKYALFDTKEMAYIP-LDQOKVKKRA-----AIDVVAR 445
Db      381 -----SKITGIMQDPKKPKGLKAPALITWKTVDKMPNVIILLGG 423
Qy      446 FKGSGGALIQGLVTCISIGAMTPY-----LAVILFTIATML-----VSATKML 493
Db      424 FALKGS-EQSGLSMLD--KLTPLOHIPESATVILCLIAITFECTSNVATTTLFLP 480
Qy      494 FLAOSA 499
Db      481 ILASMA 486

```

```

RESULT 7
US-10-014-717-11
; Sequence 11, Application US/10014717
; Publication No. US2002019278A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zitzke, Ross
; APPLICANT: Cyf, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-10-014-717-11

```

```

Query Match      4.7%; Score 122.5; DB 14; Length 713;
Best Local Similarity 21.2%; Pred. No. 0.0091;
Matches 105; Conservative 67; Mismatches 165; Indels 159; Gaps 21;

Qy      76 IIFMLIYAKLNLNLSKQ-----ALFVAVGTPLIFALFPTVYPLRDVLTHTPEPAD 127
Db      16 LALLIVTRARAGSELARRROPEVIGELRGV-----VLGFSVAGLAAGFRRALFOE 67
Qy      128 RLQAILPGLIGLVAILENMTFAFYLAELMGVMSLMFWGANETIKIHEARKFYAL 187
Db      68 PAVGVVLSGI-----SWIGALLILLMAGIEVDVIGIRKEARFAL 107
Qy      188 FGIGANISLASGRAIYVASKLRASVSEGVDPKWSLRLMLAMTIVSGLV-----LM 239
Db      108 SALGAIAPPLAAGA-----FSALVLDRLPLPSGLFGLIVSVTAIVSVIARVLIERESMR 161
Qy      240 ASYVWIN--KNVLEDPFRFYNPEEMQGGKGAPEKXNMDSFLYLDSPYIILLTLVAY 297
Db      162 RSYAQVTLAAGVSVSAWVIVAMTSSSYGASPALAVARSALL--ASGFLFPMVIV----- 215
Qy      298 GICINLIEVTV-----KSQALKIQYNNNDYSEFMGNFSFWTVGVSVLIMLFVGAVTI 349
Db      216 GRRLTHLAMRVVADATRVSKGVSVL-----VLVLTFLAALIT 252

```

```

Qy      350 RKFGM--LTGALVTPVMV-----LTGTI-----VEFALVIFRNOASGL--V 386
Db      253 QRLGHPILGAFALGVLINSAPRTNRPILDSVQLVAGLPAPVFFVLAGMVDVSQLETP 312
Qy      387 AMFGTPIMLA-----VVGAIONILSKSTKVALFSTKEMAYIPLDQOKVKKAAID 440
Db      313 AAMGTVALILLATATVAAKVPALGARLG-----GLRGEALVAVGLNM-----KGGTD 361
Qy      441 VVARFSGGALIQGLVTCISIGAMTPYLAIVLFTIA-----IMVSRATKMLKFL 495
Db      362 LIVAIVG-----VELGIL-----SNEAYTMVAVALVTVASPALLIWLE----- 401
Qy      496 AQSAKQOEVAQEDSA 511
Db      402 -----KRAPPTQESHA 412

```

```

RESULT 8
US-09-815-242-13892
; Sequence 13892, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zvekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: E11TRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13892
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13892

```

```

Query Match      4.6%; Score 120.5; DB 9; Length 395;
Best Local Similarity 20.7%; Pred. No. 0.0064;
Matches 100; Conservative 71; Mismatches 196; Indels 117; Gaps 20;

Qy      28 LPMFLMFCITFNNTVTRDKDTLIVGAPSGAAIPIFKMWLVY--PCAIIEMLIYAKL 85
Db      3 LCPMLALFALLTISAFAL-GTTEFVIAG-----LVPTIQQQLAVSIPSGMVLVIY--- 51
Qy      86 SNLSKOALFYAVGTPLIFALFPTVYPLRDVLTHTPEF--ADRLOATILPGLIGIVA 142
Db      52 -----ALGVATIAAPVLT--AL--TGRUPRKQULVALMLFTKGNLLAQAQAPGYMTLLIA 100
Qy      143 ILRNWTPAAFYVLAELMGVMSLMFWGANETIKIHEARKFYALFGIGANISILAS-- 199
Db      101 -ARLLTGLAGVVFVSISSTIATLSL-----VPEKKAASAIAMIFG-GITVALVTVGVPL 150

```

QY 200 -----GRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLIMASYMMINKNVLTPDRF 255
 DB 151 GTTICQHGWR-----ETFLAVSLGLVILMSSQLIPANT----- 186
 QY 256 YNPEMOKGKAKGAKPMNMKDSFLYLDSPYLLLTLLVIAVGICINLIEVTMSQOLKQ 315
 DB 187 -----GPRAAA-----SIRDOVKVLTTPRLLLIYAVALAGYGVF-----TAFTEL 227
 QY 316 YPNNDYSEFMGNFSEFWGV---VSVLMLFVGGVIRKFGMLTGALVTPVAVLLTGIVF 372
 DB 228 APMMODLGFSPRAVSWLLGVSVALGNWGGKLAOKHG-----AVPALKTFPALF 281
 QY 373 FALVIFRNOASGLVAFGTTPLMLAVVGAIONISKSTKYALFDPSTKEMAYIPLDORQ- 431
 DB 282 VLLMFQYLTASQYALATILVMGIFAGGNVPG-----QYVVGAKAQOF 326
 QY 432 ---KYKGAIDVVAARFGKSGGALIQOGLVIGSIGAMPYLAIVLLPIATLVSAT 488
 DB 327 TPNADVASGLNIAAFNIGIALGVSIGQTVAHYGL--AQTFWIGALIVLAVFLIMGVSG 384
 QY 489 KLNK 492
 DB 385 RLDK 388

RESULT 9
 US-09-815-242-12171
 ; Sequence 12171, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12171
 LENGTH: 425
 TYPE: PRT
 ORGANISM: *Staphylococcus aureus*
 US-09-815-242-12171

Query Match 4.6%; Score 120.5; DB 9; Length 425;
 Best Local Similarity 19.5%; Pred. No. 0.0071;
 Matches 94; Conservative 86; Mismatches 160; Indels 143; Gaps 19;
 9 FGKRSFIMPIHTHELKLKYLPMFLMFCITENYTVLRTKDTLLIYAGDAGSGAEAI----- 63
 7 FGAGNLIPPMILGHTAGCNMMWIGMLGFALT---GILLPFIIVIVAFYDEGVESVGNRIH 63

QY 64 PFIRFWLVPCALIFMLIYA-----KLSNLSKQALFYAVGT-----PFLIFPAL 108
 DB 64 PMPGPIFV---VYMGIGAFYGPRAANVA-----YEICTRILLPVHNQMTLIIIFAAV 114
 QY 109 FFTYIYPLRDVLAHTPEPADRLQAILPGLGLVAIILNWTFAAY----- 153
 DB 115 FFAIVYWS--LNSKIVDNLGKLLTPLLLMVALL---SIAVFNPEBSALSAPDKXYT 169
 QY 154 -----VLAELWGSYML-----SLMFMGFANEITKIHAKRF 184
 DB 170 HPIFSGSIEGYFTMDVLAALFVSIVYNGYFKGLTDRTKLKYVCBSGFIAAILLGMIV 229
 QY 185 YALFGIGANTS--LLASGRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLIMASY 242
 DB 230 FALAVYGASTPAGNFQKQDITLITVNSLRVFGSFGNLVFGMTV--ILACLTICIGIV-NACA 287
 QY 243 WINKNVLTPDRFPNPEMOKGKAKPMNMK-----DSLTYLDRSYIIL--LT 291
 DB 288 TFTKXHV-----PKFSYKIFALVFSIIGFLFTTLGLEMLIKIAVP 327
 QY 292 LVLVAYGICINLIEVTWKSQOLKQYPMNDYSEFMGNFSEFWGVSVLIMFVGANVIRK 351
 DB 328 LTLIYFVSIALVLSF-----ANMFSTRFSAVRLATVITLIIISIL--QILNS 375
 QY 352 FGMLTGALVTPVAVLLTGIVFALVIFRNOASGLVAFGTTPLMLAVVGAIONISKST 411
 DB 376 FNLHGVLKMKFMMPLADIDLAMLV-----PFLILAIIGTIDVIRRP 420
 QY 412 KYA 414
 DB 421 KQA 423

RESULT 10
 US-09-815-242-13831
 ; Sequence 13831, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13831
 LENGTH: 466
 TYPE: PRT
 ORGANISM: *Salmonella typhi*
 US-09-815-242-13831

Query Match 4.5%; Score 118; DB 9; Length 466;
 Best Local Similarity 20.4%; Pred. No. 0.014;
 Matches 99; Conservative 70; Mismatches 179; Indels 138; Gaps 23;

69 WLVPICAIIFMLIAKLSNI---LSKQALFYAVGTPFLIFALPEPTVIYPLRDVLPHTF 125
 41 WISVSCILLARCVMMFLSAVAVNLNKIGFDFTDQLFLL-----TAL 82
 126 ADRLOALIPGLGLVAL--RNMTPAFVYLAELMGSVMSLMWGA--NEITKHEA 181
 83 PSLGAILRVYSEFVPLFGGRKMTVLSVL-----IIPCAMLGFAVONPATP--- 131
 182 KRFPALFGICANISLASGRAIWMASKL-----RASVSEGVDPWGISTRL 227
 132 -----FQVFLILALCGFAGANFASMGNSISPPFPKAGQSALCINGLGLSVWQL 185
 228 MAMTIV-----SGLVLMASVW-WIN-KNVLDPDPFYNPEMOKKKG 267
 186 IAPVIFPIFTPLGVRGVPQPDGSLALVTNAAMITVPLAVATLAAMFGMNDIGSKAS 245
 268 AKPKNMDSFLYIDRSYIILLTLVIA-YGICINLEVTWKSQKIQYRNM-----D 321
 246 VA-----SOLPVKRLHMLSLYLATGSGFIG-ESAGFAMLAQTQPDVNILOAF 297
 322 YSEFMGNFSFWTGVSVLIMLFVGNAVIRKFGMLTGAIVTPV-NVLIGVFPALVFRN 380
 298 FGPIG-----ALASAGVDSKFGGAVVTLINIFMALFPALLFPLT---PG 343
 381 QASGLVAMF-----GTPPLMAVVGAIONILSKSTKY--ALFDSKEMA 423
 344 SGAGSFAFYLVFMGLFTAGIGSGSTFQMTAVIPRO-----TLVNVLRGSDPQA 396
 424 YIPDQEKVKGAIDVNAARFGSGGALLQGL---LVICGS-IGMPTYLAVIILFI 479
 397 -----QRAVDTDAALGFGISAGVGFIPKAGTSLATGSPVGMKIFLIFYACV 451

480 IAIMLV 485
 452 LITMLV 457

RESULT 11
 US-09-815-242-5678
 ; Sequence 5678, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR APPLICATION NUMBER: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5678
 ; LENGTH: 430
 ; TYPE: PR1
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5678

Query Match 4.5%; Score 117.5; DB 9; Length 430;
 Best Local Similarity 19.3%; Pred. No. 0.014;
 Matches 93; Conservative 86; Mismatches 159; Indels 143; Gaps 19;

9 FGKRSFLMPPIHTELKVLPMPLMPCITFENVTLBDTKOTILVAGAGSABAI----- 63
 15 FGAGNLFPPLGHTAQQNMWIGLGFALT---GILPFTIYVAYAYDESVGNRIH 71
 64 PFIFMLVPCAIIFMLIYA-----KLSNLSKQALFYAVGT-----PFLIFPAL 108
 72 PMFGFIPAV---VIYMSIGAFYGIPIRANVA-----YEIGTRHILPVHNMQWTLIIFAAV 122
 109 PFTIYVLRDVLHTEPADRLQALIPGLGLVALILNMTPAAY----- 153
 123 FFAIVYMS--LNBSKIVDMGLKLTPLLLMALL--SIAYIFNESALSAPKXYIT 177
 154 -----VLAELMGSVYL-----SIAMFGPANEITKIHAKRF 184
 178 HPFISGLEGFTMDVLAALAFSVIYNGYFKGLTTRTKILKYVCSGFIALLILGMIY 237
 185 YALFGICANIS--LLASGRAIWMASKLASVSEGVDPWGISTRLMAMTIVSGVLMASY 242
 238 FALAYGASTAPGNFKQGTDLITVNSLRVFGSGNLVFGMTV-ILACLTCTIGLV-NACA 295
 243 WIKNVNLTDRPFYNPEMOKKGAAPKXNMK-----DSLVIYDRSYIILL--LT 291
 296 TPTKGVH-----PKFSYKIFALVFSIIGLFTTLGHEMIKIAVP 335
 292 LTVIAYGICINLEVTWKSQKIQYRNMNDYSEFMGNFSFWTGVSVLIMLFVGNAVIRK 351
 336 LITLIFVSIALVLSF-----ANMFSTFRFSAIRLATVTLIISIL--QILNS 383
 352 FGMLTGALVTPVWVLTGIVFPALVIFRNOASGLVAMFGTTPMLMAVVGAIONILSKST 411
 384 FNLHGVIWKPMWMLPLADIDLAMLV-----PFLILMIGTIIIDVIVIRP 428

412 K 412
 429 K 429

RESULT 12
 US-09-815-242-13840
 ; Sequence 13840, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR APPLICATION NUMBER: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 13840
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(477)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13840

```

Query Match 4.4%; Score 116.5; DB 9; Length 477;

Best Local Similarity 20.2%; Pred. No. 0.02; Matches 95; Conservative 84; Mismatches 186; Indels 105; Gaps 27;

```

QY 69 WLVPFCATIFMLIVAKLSNI--LSKQALFYAVGTPFLIFPALPTVIYPLRDVLPTEPAD 127
DB 51 MISVCCILAPCVWMLFSAVANLPRKGFNFTDQFMILTALPSVSGALLVPPY----- 104
QY 128 RLQALIPGGLGLVALINMTFAAFYVLAELWGSVMSLMFWGFANEITKJHEAKRYAL 187
DB 105 --SKWVP--LFG---GRRWT--ARST-----GILIVPCWLGXA-----VQDSTPST 144
QY 188 FGIGANISILLASGRAIWAASKLARASVS-----EGVDPMGISRLLMATTI 232
DB 145 FII---ISLLCKXXGXGAPASSM-ANISFFPRKQGGALINGLGNNGVSVMLXAPLV 200
QY 223 V-----SGVLMAASYW-WIN-KXVLTDPFYVPEEMQKKGAKKPRM 272
DB 201 VLSLTFAPAGSHGVEQPDGSQLYLANAMWIPFLAIFTLLAFQMEL-----ATSKA 254
QY 273 NMKDSFLYLDSPYILLTLTLVIA-YGICINLIEVTWKSQKLQYPMNDYSEPMGNFSF 331
DB 255 SLKEQLPVLPKRG-HIMIMSLVLTATFGSFIG-FSAGFAMLSKTQPPDVQ-----ILHAYF 307
QY 332 WTGVSVVIMLFGVGNVIRKGMLTGALVTPY-MULLGIVFPALVITRNQASGLVAMEG 390
DB 308 FGPFGLALAR-SAGGALISDRUGTRVTLINFLVLAIFSGLLFTLPT-GGVGGSFIAPFG 365
QY 391 TTPLMLAVVV-----GAIQNLISKSTKYALFDSFKEMAYIPLDDEQKVKGAALDVAA 444
DB 366 ---VFLAFLTRAGLGSSSTFQMISVIFPKLTMDRKAGS-----GSDDEAMREAAITDAAA 418
QY 445 ----RFGSGGALIQQGL---LVICGS-IGAMTPYLAIVILLFTIAVLV 485
DB 419 LGFLSAIGAIIGFFIPKAFGSSSLALTGSPVGMKVFLLFIYACVITWAV 468

```

RESULT 13

```

US-09-738-626-3861
; Sequence 3861, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

```

; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18 19:37:484
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO: 3861
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3861

```

Query Match 4.4%; Score 115.5; DB 10; Length 410;

Best Local Similarity 22.8%; Pred. No. 0.02; Matches 72; Conservative 47; Mismatches 104; Indels 93; Gaps 16;

```

QY 157 ELKGSVWLSL--MFWGFANEITKI--HEAKRYA-----LFGIGANISLLASG-- 200
DB 20 KVMILAVLSVFTVAMG-GNEFTPLVFRGSGFNSLFDLLVFAIGYAVGLAAGPL 78
QY 201 -----RAIVWASKLRASVSGVDPMGISRLLMAM-----TIVSGVLMAASYWIMKNV 249
DB 79 SDRYGRRAVMDPALIILLSGALJASGEETAILAIARVUSGISVGNWTRAGSWIKE-- 136
QY 250 LTDRFYNPPEEMQKKGAK-PRKNMKSFL-----YLDSPYILLTLV 294
DB 137 LSSSRF---EPGVTSAGAKRASWSLTGFGALGPALAGWAQMLPLPGQALAYVHLITLL 193
QY 295 IAVGICINLIEVTWKSQKLQYPMNDYSEPMGNFSWTGVV--SVIIMLFGVGNVIRKF 352
DB 194 ILFPLIITABETROSALHKTK-----GSFWSVLYVPSALDKRFL--FVAVPI 238
QY 353 G-WLTGLVTPVWVL-----LTGIVFPALVITRNQASG----- 384
DB 239 GPWVFGAALFAYAVLPSQLDMSAPVAYSHLIALVLTLSGFGIQGPGQIMGTSKTRGP 298
QY 385 LVAMEGTTPLMLAVVV 400
DB 299 ILAMFVTVIGMIGAVI 314

```

RESULT 14

```

US-10-181-319-4
; Sequence 4, Application US/10181319
; Publication No. US20030135032A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, Thomas A.
; APPLICANT: Paszczynski, Andrzej
; APPLICANT: Crawford, Ronald L.
; APPLICANT: Cortese, Marc S.
; APPLICANT: Sebati, Jonathan L.
; TITLE OF INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: IDAHI19440
; CURRENT APPLICATION NUMBER: US/10/181,319
; CURRENT FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: PCT/US01/02386
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/177,251
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 4
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Pseudomonas stutzeri
US-10-181-319-4

```

Query Match

4.4%; Score 115; DB 12; Length 392;

Best Local Similarity 20.9%; Pred. No. 0.021;
Matches 88; Conservative 55; Mismatches 127; Indels 152; Gaps 20;

```
QY 94 LEYAGTPTLFFALFFYTYPLRDLVHTEPAD-----RLQ-AIIPGILGVALIRWT 148
DB 18 LIIAGMPWMIFFYAL-----GILGPHLVADIGISKQOLGWLTAFTGALALISPA 68
QY 149 FA-----AFVYLAELMGSVMLSLMFQFANEITKIEAKRFYALFGIGANTSL 196
DB 69 GALVQRMGRACLICFLLVGLSFLSLAVLP--GGGLVTAL-----LLCGTQOSLAN 119
QY 197 IASGRAIYWSKL-RASYSEGVDPWGISLRLLMAMTVISGLVLM-----239
DB 120 PATNQALIASVVARAKAGVGLKQSGVQASALLAGVAPLVLMGMGRGALAAVPAVALV 179
QY 240 ----ASYMINOVLTDPFNFPEKQKGAAPKPMNKDSFLYDRSPYLLLTLLTVI 295
DB 180 MAALVTYWPAAKSV-----SAPSLPLR-----VGGPNVWLSILMAI 215
QY 296 AYGICINIEVTWKSQKLQYPMNDYSBFGNFSFTGV-VSVLJMLFVGNAVIRKF-- 352
DB 216 Q--LQAGLA-----LSFMTFLGVYAAQIGSVSTI-----GAMVSCRA 253
QY 353 -GMLTGAIVTPVM-----VLTGIVF--PALVIFR--NQASGLVAFGTTPMLAVV 399
DB 254 MGILSRVLLTPIADKLKDEITLLGVLFILAGLALVMEANQGMPLMLGVTGMGLTV- 312
QY 400 VCAIWNILSKSTKYALPDSTKEMAYIPLDQEQVKGAIDVVAARFGSGGALLIQGLL 459
DB 313 --AAGNAIAMS-----MLRDGRFGGAATSAAGWL 339
QY 460 VI 461
DB 340 SV 341
```

RESULT 15

US-09-738-626-5024
; Sequence 5024, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.10
; SEQ ID NO 5024
; LENGTH: 708
; TYPE: PRT
; ORGANISM: *Corynebacterium glutamicum*
US-09-738-626-5024

Query Match 4.4%; Score 115; DB 10; Length 708;
Best Local Similarity 18.6%; Pred. No. 0.047;
Matches 106; Conservative 70; Mismatches 149; Indels 244; Gaps 25;

```
QY 109 FPTYVYPLRDLVHTEPADRIQALIPG-----LLGVAILRMTAFAYVLAEL 158
DB 120 YPHVRPAPAPAKPPTSEKIMRGVATIGGVITVAGVILLVSAIQRM-----LGPLGRV 175
QY 159 WGSVWLSLMFQFANEITKIEAKRFYALF-----GIGA 192
DB 176 IGAVYLAIVLLGAHAYVK--RGTRVEALVALLVTSQIAFLATSAIIFILEMPPGGS 233
QY 193 ----NISLASGAITYMA-SKLASVSEGVDPGISLRLLMAMTVISG-----LVMA 240
DB 234 LVALLIGNIFLIVGR--LMSLSKTEKSAEG-----HTVFGAIAVSGSAILFALSA 284
QY 241 SYMNINKVVLTDPRYNDEEMQKGAAPKPMNKDSFLYDRSPYLLLTLLTVIAYGIC 300
DB 285 DAWM-----PIESVAAALISYRIS 304
QY 301 INLI-----EYTWKSQ-----LKLQYPMNDY 322
DB 305 TIIIRASMAAPAVILQFVLSASQWQMEWPATIVGTAVLLVALLTMDPFKITATDSHI 364
QY 323 S--EPMGNF-----SFTGVSVLIMLFVGVNAVIRKFGMLTGAIVTPMVLITGVFAL 375
DB 365 ALBEYWRSPETNPVSTWGAVSPVLIIVITSMPIAVDMPWMLALIPACAVAAIGI--PAL 422
QY 376 ----VFRRQASGLVAFG-----TTPMLAV--VGA-----402
DB 423 RSSDTASTENORMSLVAVGLALIAETVQLFTGDLFTNPLVWVPLIAGALLPMLRM 482
QY 403 ----IONILSKSTKY-----ALFDSKEMAYIPLDQEQ 431
DB 483 LPQORQGVPEWVAMLIAAVAMTGVLLRNVSISPLMTDTQALQALLIVFI-----536
QY 432 KVKGAIDVVAARFGSGGALLIQGLLV--CGSIGAMPVY-----AVILFTI 480
DB 537 ----AATIQVRRSPFGHMLQIILVGLTLLTSAISIVITITFTIGRLIAGNAGMWLGRFI 592
QY 481 A-----IWLVSATKL--NKLFLAQSA 500
DB 593 GHATVSIIMWVIAAALMLNRLDLPAGAL 621
```

RESULT 16

US-09-759-143-778
; Sequence 778, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: XU, Jiangchun
; APPLICANT: DILLON, David C.
; APPLICANT: MITCHEM, Jennifer L.
; APPLICANT: HARLOCKER, Susan L.
; APPLICANT: JIANG, Yugu
; APPLICANT: HENDERSON, Robert A.
; APPLICANT: KALOS, Michael D.
; APPLICANT: FANGER, Gary R.
; APPLICANT: RETTER, Marc W.
; APPLICANT: STOLK, John A.
; APPLICANT: DAY, Craig H.
; APPLICANT: VEDVICK, Thomas S.
; APPLICANT: CARTER, Darrick
; APPLICANT: LI, Samuel
; APPLICANT: WANG, Aijun
; APPLICANT: SKEIKY, Yasir A.W.
; APPLICANT: HEPLER, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT

Qy 257 NPEEMOKGKAKPRNMKDSFLYDRSPYLLTLVIAVGCINLI----- 304
Db 841 -----SRLNGKXIMLQRM- IDVFFFLFAVMVAFGVAROGILRQNEGRMRIF 891
Qy 305 -EVTWKSQKL- -QYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRK- GMLTGAIV 360
Db 892 RSVIYEPYAMFGVPSVDVDTYDFAHCTFTGNESKPLCVELDEHNLPRPEWITIPLV 951
Qy 361 TPWVVLITGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 19

US-09-895-793-778
; Sequence 778, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-778

Query Match 4.3%; Score 113; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;
Qy 43 VLBDTKDILLVAGAGSGAEALPIKFWLVVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDRKHKULMW 717
Qy 97 AVG-----TPELI-----FPALPPTV-IYPLRDVILHPTFPADRLOALIPQLGLIVA 142
Db 718 YVAFPTSPFVFSNNVVFYIAFLLFFAYVLLMDPHSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNMTFAFYVLAELMGVMSL-MFWGFANEITRIHAKRFYALFGIGANISL 196
Db 768 VLPCDEVRQWYVNGVNYFTDLN-NVMDTLGLFYFIAGIVRRLH-----SSKSS 815
Qy 197 LASGRAIWMASKLPAVSSEGVDPWGISLRLLMAMTYISGLVLMASVWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEEMOKGKAKPRNMKDSFLYDRSPYLLTLVIAVGCINLI----- 304
Db 841 -----SRLNGKXIMLQRM- IDVFFFLFAVMVAFGVAROGILRQNEGRMRIF 891
Qy 305 -EVTWKSQKL- -QYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRK- GMLTGAIV 360
Db 892 RSVIYEPYAMFGVPSVDVDTYDFAHCTFTGNESKPLCVELDEHNLPRPEWITIPLV 951
Qy 361 TPWVVLITGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 20

US-09-895-814-778
; Sequence 778, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-778

Query Match 4.3%; Score 113; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;
Qy 43 VLBDTKDILLVAGAGSGAEALPIKFWLVVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDRKHKULMW 717
Qy 97 AVG-----TPELI-----FPALPPTV-IYPLRDVILHPTFPADRLOALIPQLGLIVA 142
Db 718 YVAFPTSPFVFSNNVVFYIAFLLFFAYVLLMDPHSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNMTFAFYVLAELMGVMSL-MFWGFANEITRIHAKRFYALFGIGANISL 196
Db 768 VLPCDEVRQWYVNGVNYFTDLN-NVMDTLGLFYFIAGIVRRLH-----SSKSS 815
Qy 197 LASGRAIWMASKLPAVSSEGVDPWGISLRLLMAMTYISGLVLMASVWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

Qy 257 NPEMOKGKGAKPMMKNSFLYDRSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKITIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRONEQRMWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKF-GMLTGAIV 360
Db 892 RSVIYEPYLFAMFGVPSVDVDTGYDFACHCTFGNESKPLCVELDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978

RESULT 21

US-10-144-678A-778
; Sequence 778, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Panger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa W.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; NUMBER OF SEQ ID NOS: 1033
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-678A-778

Query Match 4.3%; Score 113; DB 12; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13; Indels 144; Gaps 19;
Matches 85; Conservative 54; Mismatches 121;

Qy 43 VLADTKDTLIVGAPSGAEAIPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIIPLVGCGFVSFRKKVPDKHKKLWY 717
Qy 97 AVG---TPFLI-----FFALPPTV-IYPLRDVLHPTFADRLQALIPGLGLVA 142
Db 718 YVAFETSPFVFSNNVVFYIAFLFLFAVLLMDFSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNNTFEAFYVLAELMGSVMLSL-MWGFANEITKIHAKRFYALFGIGANISL 196
Db 768 VLPCEDEVQRYVNVNFTDLM-NVMDTLGLFYIAGIVFRH-----SSNKS 815
Qy 197 LASGRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLVMSYWMINKNVLTPREFY 256
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978

Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840
Qy 257 NPEMOKGKGAKPMMKNSFLYDRSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKITIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRONEQRMWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKF-GMLTGAIV 360
Db 892 RSVIYEPYLFAMFGVPSVDVDTGYDFACHCTFGNESKPLCVELDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978

RESULT 22

US-10-294-025-778
; Sequence 778, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolck, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 778
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-294-025-778

Query Match 4.3%; Score 113; DB 12; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.13; Indels 144; Gaps 19;
Matches 85; Conservative 54; Mismatches 121;

Qy 43 VLADTKDTLIVGAPSGAEAIPIKFWLVVPCALIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDYKN-----WKIILCLFIIPLVGCGFVSFRKKVPDKHKKLWY 717
Qy 97 AVG---TPFLI-----FFALPPTV-IYPLRDVLHPTFADRLQALIPGLGLVA 142
Db 718 YVAFETSPFVFSNNVVFYIAFLFLFAVLLMDFSVHPPELV-----LYSLVF 767
Qy 143 IL-----RNNTFEAFYVLAELMGSVMLSL-MWGFANEITKIHAKRFYALFGIGANISL 196
Db 768 VLPCEDEVQRYVNVNFTDLM-NVMDTLGLFYIAGIVFRH-----SSNKS 815
Qy 197 LASGRAIWMASKLRASVSEGVDPWGISLRLLMAMTIVSGVLVMSYWMINKNVLTPREFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840
Qy 257 NPEMOKGKGAKPMMKNSFLYDRSPYLLTLTLVAVGICINLI----- 304
Db 841 -----SRNLGPKITIMLQRM-L-IDVFFFLFAVMMVAFGAROGILRONEQRMWIF 891
Qy 305 -EVTWKSQKL--QYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKF-GMLTGAIV 360
Db 892 RSVIYEPYLFAMFGVPSVDVDTGYDFACHCTFGNESKPLCVELDEHNLPRPEWITIPLV 951
Qy 361 TPVWVLLTGIVFPAIVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978

RESULT 23
US-10-012-896-778
; Sequence 778, Application US/10012896
; Publication No. US20020183251A1

Sequence 11, Application US/10171319
Publication No. US20030157633A1
GENERAL INFORMATION:
APPLICANT: Arden Patapoutian
APPLICANT: Andrea Peier
APPLICANT: Peter McIntyre
APPLICANT: Stuart Bevan
APPLICANT: Chuansheng Song
APPLICANT: Pamosh Ganju
TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
FILE REFERENCE: 4-32048A
CURRENT APPLICATION NUMBER: US/10/171,319
CURRENT FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: 60/297,835
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: 60/351,238
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: 60/352,914
PRIOR FILING DATE: 2002-01-29
PRIOR APPLICATION NUMBER: 60/357,161
PRIOR FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: 60/381,086
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: 60/381,739
PRIOR FILING DATE: 2002-05-16
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 1268
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-319-11

Query Match 4.3%; Score 113; DB 12; Length 1268;
Best Local Similarity 21.0%; Pred. No. 0.16;
Matches 85; Conservative 54; Mismatches 121; Indels 144; Gaps 19;

QY 43 VLRTKOTLIVGAGSGAEALPFIKMLVPCALIFMLI-----YAKLSNLSKQALFY 96
DB 850 ISRDYKN-----WKIILCLFPIIPVGGFVSFRKKPYDKHKKLWY 890
QY 97 AVG---TPFLI-----FFALFPTV-IYPLADVLHPTFEADRLQAILPGLGLVA 142
DB 891 YVAFPTSPFVPSNVVYIYFLLFAYVLLMDFSVHPBELV-----LYSLVF 940
QY 143 IL-----RNWTFAPFYVLAELMGSVYMLSL-MFWGFANETIKIHEAKRYVALFGIGANISL 196
DB 941 VLFCEVQWYVNGVNYFTDLW-NVMDTLGLFYFAGIVPRHL-----SSNKS 988
QY 197 LASGAIYVASKLRASVSEGVDPWGISIRFLMAMTIVSGVLVLMASYMINKVLTDPREFY 256
DB 989 LYSGRVIFC-----LDYIIFTLRLIHFIV----- 1013
QY 257 NPEEMQKKKAKPKMNKDSFLYDRSPYILLTLVIAVGICINLI----- 304
DB 1014 -----SRNIGPKITIMLQRM-LIDVFFFLFPAVMAAGVARGILRQNEGRWTF 1064
QY 305 -EVTWKSQKL--QYNNANDYSEFWGNFSFWTGVVSLIMLFGVGNVARKF-GWLTGALV 360
DB 1065 RSVIYEPYLFAMFGQVPSVDGTTVPFACHTFGNSKPLCYELDEHNLPRPEWTTIPLV 1124
QY 361 TPVWVLLTGIVFPALVIRNOASGVAMFGTTPLMLAVVGAIQ 404
DB 1125 C-IYVLTSTNIIILVNL-----LVAMFGYT-----VGTGV 1151

RESULT 26
US-09-920-653-3
Sequence 3, Application US/09920653
Patent No. US20020104113A1
GENERAL INFORMATION:
APPLICANT: Japan as Represented by Director General of Okazaki National Research

APPLICANT: Institues
TITLE OF INVENTION: Nav2 channel gene-deficient non-human animals
FILE REFERENCE: J2001P059
CURRENT APPLICATION NUMBER: US/09/920,653
CURRENT FILING DATE: 2001-10-16
PRIOR APPLICATION NUMBER: JP 2000/237320
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: JP 2000/241637
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: JP 2001/222263
PRIOR FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 1681
TYPE: PRT
ORGANISM: Mus musculus
US-09-920-653-3

Query Match 4.3%; Score 113; DB 10; Length 1681;
Best Local Similarity 21.4%; Pred. No. 0.23; Indels 106; Gaps 18;
Matches 78; Conservative 56; Mismatches 124;

QY 3 KTEKPKGKLSFPLMPTHTHELKKVLPMLMFCITENYTVLRDTKOTLIVGAGSGAEA 62
DB 1231 KPTPRPNKKGGLFLDLVTHRVFVILIL-----ICFOATTIMIQD-----EQ 1275
QY 63 IPIFK--FWL-----VPCAI-----IMLYAKLSNLSKQAL 94
DB 1276 SPQMETAIIPWNSIFVMLFTLECLIKLAPRCHYFSAVMNVHDVVIFSTGLLPLTI 1335
QY 95 FYAVGTPELIFPALFPTVYVPLRDVLHPTFEADRLQAILPGLGLVAIRNWTFAFYV 154
DB 1336 GQIVPSPVLQVLLISVYIHLRPGKPKVPHD-----LMDPLIALALL-NISLLIFLV 1390
QY 155 LAELMGSVMLMFWGFANETIKIHEAKRYVALFGIGANISLASGAIYVASKLRASVS 214
DB 1391 -----NW-----IYALFGM-YNFAYV-----KKEAGIN 1412
QY 215 E-GVDWGISIRLLMAMTIVSG---LVLMASYMINKVLTDPREFYNEPEMKKKGA 268
DB 1413 DVSNFETFGSSMLCLFQVTTFSGWDMLDIAFNSQWSD---CPDRIINPQYKGCOS 1468
QY 269 KPKNMKDSFLYDRSPYILLTLVTAAGICINLIVTWKSQKLQYNNANDYSEFWGN 328
DB 1469 -PSVGT--SYFV---SYLISWLIYVMTIVLIMEFLSLPSQKSGRTLSDDPRRF--- 1518
QY 329 FSPW 332
DB 1519 FRVW 1522

RESULT 27
US-09-738-626-6914
Sequence 6914, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16

/ PRIOR APPLICATION NUMBER: JP 00/159162
 / PRIOR FILING DATE: 2000-04-07
 / PRIOR APPLICATION NUMBER: JP 00/280988
 / PRIOR FILING DATE: 2000-08-03
 / NUMBER OF SEQ ID NOS: 7059
 / SOFTWARE: PatentIn ver. 3.0
 / SEQ ID NO: 6914
 / LENGTH: 1083
 / TYPE: PRF
 / ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6914

Query Match 4.2%; Score 111.5; DB 10; Length 1083;
 Best Local Similarity 19.0%; Pred. No. 0.18;
 Matches 114; Conservative 80; Mismatches 170; Indels 237; Gaps 28;

QY 12 LRSFLPIITHELKVLPMFLMF--CITFNVTLDYDTDLIVAGPSGAEAIPIKFW 69
 DB 105 LLSLVIPVLTAKEDADGSGFFRLTLSTVLGGVTLISIGAP----- 151
 QY 70 LVVPCAIIFMLYAKLS-----NLSKALFYAVGTPFLIFALPTVYPL--RDVLRP 122
 DB 152 -----LITRMMLSSGQVNVWSTAPAYWL-LPQIFFGFLFMAVINTREVEKP 201
 QY 123 TEFAURLQALPPLGLGLVAIIRMTFAAFVYLAEIMGSVMSLMFWGPAEITKHEAK 182
 DB 202 GAMAPVNNVITLTVLGVVWLP-----ARLHPHE 231
 QY 183 RF-----YALFGIGANISLASGRAIYVASKIRASVSEGVDP--WGISLRLL--MA 229
 DB 232 QVGIRDPQIIFLGVGTTLGVA--OCLIMIPLYRRA--GIDMRPLMGIDARLKQFGWA 286
 QY 230 MTIV-----SGLVMASYWMINKVLD--PRFN 257
 DB 287 NAIIVYVAISQFGYIITRIASIAIDADAPFIYQOHHMLQVPGIIGVTLTAIMEPLSR 346
 QY 258 -----PEMOKGKKAKPKNMKDSPLYLRSFYILLTLTLLVI--AYGICINILE 305
 DB 347 NAADDDBRAVSDLLQSGK-----LTFIALIPVYFTAGV----- 383
 QY 306 VTMKSQKLQYNNND--YSEFMGN-----FSFWGVSVSLIMLFGVGNVIRKP 352
 DB 384 -----PIANGLFAYVGFDPANAANILGWTLSFSFAFTLIPYALVTLHLRVEFARBE 432
 QY 353 GMLTGALVTPVWVLLTGIVFALVIFRNASGLVAMEGTPMLAVVVAIONI----- 406
 DB 433 VW-----TPFTI-TAGITATKVL-----SLAPLLSSPERVVVLLGANGFSFITGA 480
 QY 407 -----LSKSTKXALFDSTKEMAVIPLDQEKVKGAIDVVA--RFG 447
 DB 481 VIGAVLNNKGLGLGMRSLAKTSLWAL-----GSAVGAAMAAALG 521
 QY 448 KSGGALIOGGLIVIGSIGAMTPYLAIVILLFTIIMVLSATKLKFLAQSALKE-QEVA 506
 DB 522 WLIAQVGDPLGLTSSVGYLL-YLAVLGVFIFV-----TGIVLSRSGLPVQNLG 572
 QY 507 Q 507
 DB 573 Q 573

RESULT 28
 US-09-738-626-4455
 / Sequence 4455, Application US/09738626
 / Publication No. US20020197605A1
 / GENERAL INFORMATION:
 / APPLICANT: NAKAGAWA, SATOSHI
 / APPLICANT: MIZOGUCHI, HIROSHI
 / APPLICANT: ANDO, SEIKO
 / APPLICANT: HAYASHI, MIKIRO
 / APPLICANT: OCHIAI, KEIKO
 / APPLICANT: YOKOI, HARUHIKO
 / APPLICANT: TATEISHI, NAOKO

/ APPLICANT: SENOH, AKIHIRO
 / APPLICANT: IKEDA, MASATO
 / APPLICANT: OZAKI, AKIO
 / TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 / FILE REFERENCE: 249-125
 / CURRENT APPLICATION NUMBER: US/09/738,626
 / CURRENT FILING DATE: 2000-12-18
 / PRIOR APPLICATION NUMBER: JP 99/377484
 / PRIOR FILING DATE: 1999-12-16
 / PRIOR APPLICATION NUMBER: JP 00/159162
 / PRIOR FILING DATE: 2000-04-07
 / PRIOR APPLICATION NUMBER: JP 00/280988
 / PRIOR FILING DATE: 2000-08-03
 / NUMBER OF SEQ ID NOS: 7059
 / SOFTWARE: PatentIn ver. 3.0
 / SEQ ID NO: 4455
 / LENGTH: 741
 / TYPE: PRF
 / ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4455

Query Match 4.2%; Score 111; DB 10; Length 741;
 Best Local Similarity 18.9%; Pred. No. 0.12;
 Matches 108; Conservative 83; Mismatches 205; Indels 176; Gaps 24;

QY 27 VLPMLMFECIT-----FNVTLDYDTDLIVAGPSGAEAIPIKFWLV 72
 DB 204 LIPSTIMGSLSTIIAVALMLISPEVFTISASRQRTPALASQ-----APPRIHMAVL 259
 QY 73 PCALIFMLIYAKLSNLSKQALF-----YAVGTPFLI--FFAL-----FP 110
 DB 260 TYGLFAGLVGASIGLVLOIGIYGWKKTYPEFSITLTVLVGVWALAIASITAAFLP 319
 QY 111 TYIPLRDVLRHT--EPADRLOALIPGLGLVAIIRMTFAAFVYLAEIMGSVMSLMF 168
 DB 320 AVFVSRSSITINGIYIGSIDKIIIRWSPRMLIGIYVIAAVALAFIGDGE-WGVVQKQCF 378
 QY 169 WGFANEIRKIHAKFVLPFGIGANISLASGRAIYV-----SKURASVS 214
 DB 379 -----IAAVIALPASVPAVLMALGRLGLTFKALTRDMLRRSN- 416
 QY 215 EGVDPWGISLRILNMTIV--SGLVMASYWMINKVLDPRFYNEEMOKGKKAKPKM 272
 DB 417 HSIPALGALVAVIMLGTMTGATQASDEATASVPEAVFLRGDTQIPIGMOKIDV 476
 QY 273 ---NNKDSFLYD---RSPYILLTLL---VIAYGICINILEVWKSQKLQYPN--- 318
 DB 477 YGDHNGFGIYELDVDFYSANYVPALTSFPGFPVIAIPKILDMFGV--HEQADIVAPSTYN 534
 QY 319 ---NDYSEFMGNFSFWTGVSVLMLF-----VGVNVIRKFGMLT--GALVTP----- 362
 DB 535 SGLQRYAIVPDDEYMLDPAVLPPLVSHVLSPEPFEIIGQTEFLDTIVLPOLBDQT 594
 QY 363 -----VMVLLTGIVFALVIF-----RNASGLVAMEGTP- 393
 DB 595 VQAINRSDAFPSHDGHNSSLAASALTRAIVVSVLIVANRKLQOHALAL- GATPG 653
 QY 394 -----LMLAVVGAIONILSKSTKXALFDSTKEMAVIPLDQEKVKGAIDVVA 444
 DB 654 TIYKVNALNALMLLVGIMGVSGWIALLTGTDEI-----VDGAIL 697
 QY 445 RFGKSGGALIOGGLIV-----ICSGTGM 468
 DB 698 NYGTLEHMLPWLPLVLSLVVAPLVCAGVIGAI 729

RESULT 29
 US-09-759-143-780
 / Sequence 780, Application US/09759143
 / Patent No. US20020022248A1
 / GENERAL INFORMATION:
 / APPLICANT: Xu, Jiaengchun
 / APPLICANT: Dillon, Davin C.

```
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yuqi
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aljun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Hepler, William
/ APPLICANT: Houghon, Raymond L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C23
/ CURRENT APPLICATION NUMBER: US/09/759,143
/ CURRENT FILING DATE: 2001-01-12
/ NUMBER OF SEQ ID NOS: 934
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 780
/ LENGTH: 1095
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (1)...(1095)
/ OTHER INFORMATION: Xaa = Any Amino Acid
US-09-759-143-780
```

Query Match 4.2%; Score 111; DB 9; Length 1095;

Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;

Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

```
QY 43 VLDRKDTLIVGAPSGAIAIPFIKFWLVPQAIIFMLI-----YAKLSNLSKQALFY 96
DB 677 ISRDTRN-----WKIILCLFIPLVGGCFVSFRKKPVDKHKLLMY 717
QY 97 AVG---TPFLI-----FFALFPTV-IYPLRDVLHPTFADRLQALPPGLGLVA 142
DB 718 YVAFPTSPFVFSMNWVFYIAFLLFAVYVLMDFHSVHPPELV-----LYSLVF 767
QY 143 IL-----RNWTPAFAFYLAELMGVSL-MFWGFANEITIKHEAKRYALFGIGANISL 196
DB 768 VLFCEVQRQWYVNGVNYFTDLN-NVMDTGLFYFIAGIVFRILH-----SSNKS 815
QY 197 LASGRAIYVASKLRASVSEGVDPWGISRLMLAMTIVSGVLMSYWMINKVLTDPREF 256
DB 816 LYSGRVIFC-----LDYIIFTRLHIHFTV-----LVSIVF 840
QY 257 NPEEMQKKGAKPKPMNKDSFLYLDSPYIILLTLVIAIGICINLI----- 304
DB 841 -----SRNLGPKTIMLQRMIL-IDVFFFLFELFAWMVAFGVARQGLRQNEQRWIF 891
QY 305 -EYTKSQQLK--QYPNNDYSEFGNFSFMTGVSVILMLFVGGVIRKR-GLTGLALV 360
DB 892 RSVIIEPYLAMFGVPSDVTDTYDFAHCTFGNESKPLCVSLDEHNLPREFEWITIPLV 951
QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMEGTTPLMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978
```

RESULT 30

US-09-780-669-780

; Sequence 780; Application US/09780669

; Patent No. US20020051977A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, David C.

; APPLICANT: Mitcham, Jennifer L.

```
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yuqi
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aljun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Hepler, William
/ APPLICANT: Hural, John
/ APPLICANT: McNeill, Patricia D.
/ APPLICANT: Houghon, Raymond L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C24
/ CURRENT APPLICATION NUMBER: US/09/780,669
/ CURRENT FILING DATE: 2001-02-09
/ NUMBER OF SEQ ID NOS: 943
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 780
/ LENGTH: 1095
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (1)...(1095)
/ OTHER INFORMATION: Xaa = Any Amino Acid
US-09-780-669-780
```

Query Match 4.2%; Score 111; DB 9; Length 1095;

Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;

Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

```
QY 43 VLDRKDTLIVGAPSGAIAIPFIKFWLVPQAIIFMLI-----YAKLSNLSKQALFY 96
DB 677 ISRDTRN-----WKIILCLFIPLVGGCFVSFRKKPVDKHKLLMY 717
QY 97 AVG---TPFLI-----FFALFPTV-IYPLRDVLHPTFADRLQALPPGLGLVA 142
DB 718 YVAFPTSPFVFSMNWVFYIAFLLFAVYVLMDFHSVHPPELV-----LYSLVF 767
QY 143 IL-----RNWTPAFAFYLAELMGVSL-MFWGFANEITIKHEAKRYALFGIGANISL 196
DB 768 VLFCEVQRQWYVNGVNYFTDLN-NVMDTGLFYFIAGIVFRILH-----SSNKS 815
QY 197 LASGRAIYVASKLRASVSEGVDPWGISRLMLAMTIVSGVLMSYWMINKVLTDPREF 256
DB 816 LYSGRVIFC-----LDYIIFTRLHIHFTV-----LVSIVF 840
QY 257 NPEEMQKKGAKPKPMNKDSFLYLDSPYIILLTLVIAIGICINLI----- 304
DB 841 -----SRNLGPKTIMLQRMIL-IDVFFFLFELFAWMVAFGVARQGLRQNEQRWIF 891
QY 305 -EYTKSQQLK--QYPNNDYSEFGNFSFMTGVSVILMLFVGGVIRKR-GLTGLALV 360
DB 892 RSVIIEPYLAMFGVPSDVTDTYDFAHCTFGNESKPLCVSLDEHNLPREFEWITIPLV 951
QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMEGTTPLMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVO 978
```

RESULT 31

US-09-822-827-780

; Sequence 780; Application US/09822827

; Patent No. US20020081680A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C1
CURRENT APPLICATION NUMBER: US/09/822.827
CURRENT FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1095)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-822-827-780

Query Match 4.2%; Score 111; DB 9; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2;
Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

QY 43 VLPRTKDTLIVGARGSGAEALPFIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKILMY 717

QY 97 AVG---TPFLI-----FPALPPTV-IYPLRDVLPTEFADRLQALIPGLGLGVA 142
Db 718 YVAFTSPFVFSNNVFIYAFLLFAYVLLMDPHSHPHPELV-----LYSLVF 767

QY 143 IL-----RNMTFAFYVLAELMGSVMSL-MFWGFANEITKIHEAKFYALFGIGANISL 196
Db 768 VLFCDDEVQWYVNGVNYFTDLM-NVMDTLGLFYIAGIVPRHL-----SSNKSS 815

QY 197 LASGRAIYVMSKLRASVSEGVDPWGISLRLLMAMTIYSGVLVMSYMWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

QY 257 NPEEMQKKKAKPRKMKDSFLYIDRSFYILLTLTLLVIAVGCINLI----- 304
Db 841 -----SRNLGPKIIMLQRL-IDVFFFLFPAKMMVAFGVARQGLLRONEQRWRWIF 891

QY 305 -EYTWKSQQLK--QYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKF-GMLTGALV 360
Db 892 RSVIYBYPLAMFGQVPSVDGTTYPACHTFTGNSKPLCVBLDEHNLPRPEWITIPLV 951

QY 361 TPVAVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTQV 978

RESULT 32
US-09-895-793-780
Sequence 780, Application US/09895793
Publication No. US20020192763A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.

APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C2
CURRENT APPLICATION NUMBER: US/09/895.793
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1095)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-895-793-780

Query Match 4.2%; Score 111; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2;
Matches 85; Conservative 53; Mismatches 122; Indels 144; Gaps 19;

QY 43 VLPRTKDTLIVGARGSGAEALPFIKFWLVPCAIIFMLI-----YAKLSNLSKQALFY 96
Db 677 ISRDTKN-----WKIILCLFIPLVCGGFVSFRKKPVDKHKKILMY 717

QY 97 AVG---TPFLI-----FPALPPTV-IYPLRDVLPTEFADRLQALIPGLGLGVA 142
Db 718 YVAFTSPFVFSNNVFIYAFLLFAYVLLMDPHSHPHPELV-----LYSLVF 767

QY 143 IL-----RNMTFAFYVLAELMGSVMSL-MFWGFANEITKIHEAKFYALFGIGANISL 196
Db 768 VLFCDDEVQWYVNGVNYFTDLM-NVMDTLGLFYIAGIVPRHL-----SSNKSS 815

QY 197 LASGRAIYVMSKLRASVSEGVDPWGISLRLLMAMTIYSGVLVMSYMWINKNVLTDPRFY 256
Db 816 LYSGRVIFC-----LDYIIFTLRLIHIFTV----- 840

QY 257 NPEEMQKKKAKPRKMKDSFLYIDRSFYILLTLTLLVIAVGCINLI----- 304
Db 841 -----SRNLGPKIIMLQRL-IDVFFFLFPAKMMVAFGVARQGLLRONEQRWRWIF 891

QY 305 -EYTWKSQQLK--QYPMNDYSEFMGNFSFWTGVSVLIMLVGQNVIRKF-GMLTGALV 360
Db 892 RSVIYBYPLAMFGQVPSVDGTTYPACHTFTGNSKPLCVBLDEHNLPRPEWITIPLV 951

QY 361 TPVAVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIG 404
Db 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTQV 978

RESULT 33
US-09-895-814-780
Sequence 780, Application US/09895814
Publication No. US20020193296A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Heppler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 990
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1095)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-895-814-780

Query Match 4.2%; Score 111; DB 10; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;
Matches 85; Conservative 53; Mismatches 122;

43 VLKPTKTLIVGAPSGAEAIPIKFWLVPCALIFMLI-----YAKLSNLSKQALFY 96
677 ISRTKKN-----WKIILCLFIPLVCGCFVSFRKKPVDRKHKGLMY 717
QY 97 AVG---TFPLI-----FFALPPTV-IYPLRDVLPHTFADRLQAILPGLGLVA 142
DB 718 YVAFSTSPFVFSNVVFFIYAFLLFAYVLLMDPSVHPPELV-----LYSLVF 767
QY 143 IL-----RNMTFAFYVLAELMGSVMSL-MFWGFANITKIKHEAKRYALFGICANISL 196
DB 768 VLFCEVRYQWYVNGVNYFTDLN-NVMDTLGFTFIAGIVFRHL-----SSNKS 815
QY 197 LASGRAIWMASKLRASVSEGVDPWGISRLMAMTIIVSGLVMSYWMINKNVLTDPREFY 256
DB 816 LYSGRVIFC-----LDYIIFTLRLHIHFTV----- 840
QY 257 NPEMQKKKGAKEPKNMKDSFLYDRSPYIILLTLVIVAGICINLI----- 304
DB 841 -----SRNLGPKIIMLQRMIL-IDVFFFLFLFAKMWVAFGARQGLRONEQRWEIF 891
QY 305 -EYTWKSQTL--QYPMNDYSEFMNGFSFMTGVSVLIMLFVGNAVIRKF-GMLTGALV 360
DB 892 RSVIYEPYLAIFGVPSVDGTYDPAHCTFTGNEKFLCVELDEHNLPRPEWITPLV 951
QY 361 TPVWVLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 34
US-10-144-678A-780
Sequence 780, Application US/10144678A
Publication No. US20030157089A1
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Heppler, William T.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals y de Bassols, Carlota
APPLICANT: Foy, Teresa M.
APPLICANT: Matanabe, Yoshihiro
APPLICANT: Deng, Ta
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C28
CURRENT APPLICATION NUMBER: US/10/144,678A
CURRENT FILING DATE: 2002-08-12
NUMBER OF SEQ ID NOS: 1033
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 780
LENGTH: 1095
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: 867
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-144-678A-780

Query Match 4.2%; Score 111; DB 12; Length 1095;
Best Local Similarity 21.0%; Pred. No. 0.2; Indels 144; Gaps 19;
Matches 85; Conservative 53; Mismatches 122;

43 VLKPTKTLIVGAPSGAEAIPIKFWLVPCALIFMLI-----YAKLSNLSKQALFY 96
677 ISRTKKN-----WKIILCLFIPLVCGCFVSFRKKPVDRKHKGLMY 717
QY 97 AVG---TFPLI-----FFALPPTV-IYPLRDVLPHTFADRLQAILPGLGLVA 142
DB 718 YVAFSTSPFVFSNVVFFIYAFLLFAYVLLMDPSVHPPELV-----LYSLVF 767
QY 143 IL-----RNMTFAFYVLAELMGSVMSL-MFWGFANITKIKHEAKRYALFGICANISL 196
DB 768 VLFCEVRYQWYVNGVNYFTDLN-NVMDTLGFTFIAGIVFRHL-----SSNKS 815
QY 197 LASGRAIWMASKLRASVSEGVDPWGISRLMAMTIIVSGLVMSYWMINKNVLTDPREFY 256
DB 816 LYSGRVIFC-----LDYIIFTLRLHIHFTV----- 840
QY 257 NPEMQKKKGAKEPKNMKDSFLYDRSPYIILLTLVIVAGICINLI----- 304
DB 841 -----SRNLGPKIIMLQRMIL-IDVFFFLFLFAKMWVAFGARQGLRONEQRWEIF 891
QY 305 -EYTWKSQTL--QYPMNDYSEFMNGFSFMTGVSVLIMLFVGNAVIRKF-GMLTGALV 360
DB 892 RSVIYEPYLAIFGVPSVDGTYDPAHCTFTGNEKFLCVELDEHNLPRPEWITPLV 951
QY 361 TPVWVLTGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIQ 404
DB 952 C-IYMLSTNILLVNL-----LVAMFGYT-----VGTVQ 978

RESULT 35
US-10-294-025-780
Sequence 780, Application US/10294025
Publication No. US20030185830A1
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
APPLICANT: Stolk, John A.
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND


```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10467
; LENGTH: 425
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-815-242-10467
```

```

Query Match      4.2%; Score 110.5; DB 9; Length 425;
Best Local Similarity 19.7%; Pred. No. 0.063;
Matches 93; Conservative 68; Mismatches 139; Indels 173; Gaps 24;
```

```

QY      KLSNLSKQALFPA-----VGTPELFFALFPYIVLRDVLPHTPEFADRLQALIPGL 137
DB      27 KQVNPQKRALFSAMLGIVDFDPMMIF-----YLIHIKADLGITD---IQAT---L 74
QY      138 LGLVALIRNMTFAAFY-VLAELMGSVMLSLMFWG-FANEITKIHAKRFYALFGIGANIS 195
DB      75 IGTVAFIARPIGGFFGAMADKYG--RKPMMMAIIFYSGT-----GLSGIATNLY 124
QY      196 LLAGRAIV--WASKLRASVSEGVDPW-----GISRLIMAMTIVS--- 234
DB      125 MLAVCRFIVGLGMSGEYACASTYAVESWPKNLQSKASAFVSGSVGNIIAQQIIPQFAE 184
QY      235 -----GLVMASTWINKNVLTDRFPNPEEMQKKGKAKPKKMMKDSFLYLDRS 284
DB      185 VYGMNSFFIGLLPVLVLMIRKSA-----PE---SOEWIEDKXDKSTFLSVFRK 232
QY      285 PYILLTLTLVIAVIGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIML-- 342
DB      233 PH---LSISMIVFLVCFCLFGANW-----PINDLPSYLDANGVNTVISTLMTING 281
QY      343 -----FVGNVIRKFGMLTGAIVTPVMVL-----LTGIVFPAIVFR 379
DB      282 LGLTNGTIFFGFVDKIGVKKAFVVG-LITSFIFLCPFFISVKNSSILGLCLFGIMFTN 340
QY      380 NQASGLVAMFGTTPLMLAVVGAIONILSKSTKXALPDSTKEMAYIPLDQEQYKXKAI 439
DB      341 LGIAGLVPRF-----IYD-----YFP-----TKLUG--- 361
QY      440 DVVAARFGSGGALIQOGLLVIGSIGAM-TPYLAV-----ILLFIIAI 482
DB      362 -----LGTGLIYNLGATGMAAPVLAITYISGYGGLSVLFIYTV 400
```

```

RESULT 38
US-10-287-274-428
; Sequence 428, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF
; FILE REFERENCE: ELIIRA.0080V1
; CURRENT APPLICATION NUMBER: US/10/287,274
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
```

```

; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428
; LENGTH: 425
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-287-274-428
```

```

Query Match      4.2%; Score 110.5; DB 12; Length 425;
Best Local Similarity 19.7%; Pred. No. 0.063;
Matches 93; Conservative 68; Mismatches 139; Indels 173; Gaps 24;
```

```

QY      84 KLSNLSKQALFPA-----VGTPELFFALFPYIVLRDVLPHTPEFADRLQALIPGL 137
DB      27 KQVNPQKRALFSAMLGIVDFDPMMIF-----YLIHIKADLGITD---IQAT---L 74
QY      138 LGLVALIRNMTFAAFY-VLAELMGSVMLSLMFWG-FANEITKIHAKRFYALFGIGANIS 195
DB      75 IGTVAFIARPIGGFFGAMADKYG--RKPMMMAIIFYSGT-----GLSGIATNLY 124
QY      196 LLAGRAIV--WASKLRASVSEGVDPW-----GISRLIMAMTIVS--- 234
DB      125 MLAVCRFIVGLGMSGEYACASTYAVESWPKNLQSKASAFVSGSVGNIIAQQIIPQFAE 184
QY      235 -----GLVMASTWINKNVLTDRFPNPEEMQKKGKAKPKKMMKDSFLYLDRS 284
DB      185 VYGMNSFFIGLLPVLVLMIRKSA-----PE---SOEWIEDKXDKSTFLSVFRK 232
QY      285 PYILLTLTLVIAVIGICINLIEVTWKSQKLQYPMNDYSEFMGNFSFMTGVSVLIML-- 342
DB      233 PH---LSISMIVFLVCFCLFGANW-----PINDLPSYLDANGVNTVISTLMTING 281
QY      343 -----FVGNVIRKFGMLTGAIVTPVMVL-----LTGIVFPAIVFR 379
DB      282 LGLTNGTIFFGFVDKIGVKKAFVVG-LITSFIFLCPFFISVKNSSILGLCLFGIMFTN 340
QY      380 NQASGLVAMFGTTPLMLAVVGAIONILSKSTKXALPDSTKEMAYIPLDQEQYKXKAI 439
DB      341 LGIAGLVPRF-----IYD-----YFP-----TKLUG--- 361
QY      440 DVVAARFGSGGALIQOGLLVIGSIGAM-TPYLAV-----ILLFIIAI 482
DB      362 -----LGTGLIYNLGATGMAAPVLAITYISGYGGLSVLFIYTV 400
```

```

RESULT 39
US-09-738-626-5741
; Sequence 5741, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
```

SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 5741
 LENGTH: 476
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-5741

Query Match 4.2%; Score 110.5; DB 10; Length 476;
 Best Local Similarity 21.6%; Pred. No. 0.073; Indels 79; Gaps 14;
 Matches 55; Conservative 44; Mismatches 77;

QY 209 LRASVSECVD-----PWGISRLMAMTIVSGVLMAASYWINKNVLTDRFYNPE--E 260
 DB 205 LIASLASQMDVNIPTWNAVG-----AIVGVALIVPMVYKI-----YPELKD 251
 QY 261 MOKGKGAKKPMKMDSFYIDR---SPYIILTL-----LVIA-----YGCIMLI 304
 DB 252 TPEVKQASDRIKQIGFTYGEKVLGTFVVLILMTGDLVLGISATTTFVGVIIILV 311
 QY 305 E--VTWKS--OLKIQPMNNDYSEFMGNFSFMTGVSVLIMLVFGANVIRKFG--WLTG 357
 DB 312 AHVLTWEDIQEKJAMDITM-----VMPAV-----LYMMATALSQYGFIAWIS 354
 QY 358 ALVTPV-----MVLITGIVFPALVIFRNOASGLVAMF-----GTPPLMLAVV 399
 DB 355 VIASSIGGMNVVALVVLVLFPSHYFASATAHISMYLAFLGAALATGAPPLMALV 414
 QY 400 VQALQNIISKSTKYA 414
 DB 415 LAYTSNLFSSLTQYS 429

RESULT 40

US-10-032-585-7923
 Sequence 7923, Application US/10032585
 Publication No. US20030180953A1
 GENERAL INFORMATION:
 APPLICANT: Terry, Roemer D.
 APPLICANT: Bo, Jiang
 APPLICANT: Charles, Boone
 APPLICANT: Howard, Bussey
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 FILE REFERENCE: 10182-005-999
 CURRENT APPLICATION NUMBER: US/10/032,585
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 8000
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 7923
 LENGTH: 927
 TYPE: PRT
 ORGANISM: Candida albicans
 US-10-032-585-7923

Query Match 4.2%; Score 110; DB 12; Length 927;
 Best Local Similarity 20.9%; Pred. No. 0.2; Indels 160; Gaps 24;
 Matches 90; Conservative 60; Mismatches 121;

QY 66 IKFPLVPCALIFMLIYAKLSNI--LSKQALFYAVGTPFLIFALPPTVIYPLRDVILHPT 123
 DB 397 IIFW-CIPVAV-----GAISINIVLTDR-----VFELKFIILKMPDIVIMGVITGLLPV 443
 QY 124 EFADRLGAILPGLGLVALIRNWTFAFYVLAELMGVMLS-----LMFGFANEITKI 178
 DB 444 VALAILMSLVPPFI-----KM-----MGKISGRLTIQVESYCQSYFAFQVNVN 488
 QY 179 HEARFYALFGIGANISILAS-----GRAIVMASKLRASVSEGVDPWGISRLIMAMTI 232
 DB 489 -----FLAIALGSSAAVAATQIVONPEAL--QKLSSEFPKSVN--FYIYLCLEGITI 538
 QY 233 VSGVLAM-----ASYWINKNVLTDRFYNPEEMOKGKKGAAPKRNMK 275
 DB 539 SSGVLTQIVALLSHILGRIIDGTIPRAKWTMTLIGQPAY----- 578

QY 276 DSFLYDRSPYIILTLTLVIAVGICINLIEVTWKSQKQGYPMNNDYSEFMGNFSFMTGV 335
 DB 579 -STIX-----PQFQTLITVIALSYVIAPLI-----LG-----FTAI 608
 QY 336 VSVLI-----MLFV-----GGNVIRK--FGWLTGALVTPVWVLTGIVEFA----- 374
 DB 609 AFILFYFAYIYTMIFVLRPSTVDARGTNVYKSLFQFTGLFLAQLM--TALFVFSKWA 666
 QY 375 -----LVIFRNOASGLVAMFGTTPMLAVVGAIONILSKSTKYALPDSIKEMAYIPLD 428
 DB 667 CVALEGVIVVVTIARLMMKKFLPLVDVAVISAI-----KTAAGDPT--YSYPIHD 716
 QY 429 Q---EOKVKGK 436
 DB 717 QGLKEIVBGR 727

Search completed: November 25, 2003, 10:21:55
 Job time : 59 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:11:17 ; Search time 64 Seconds
(without alignments)
2076.517 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2530
Sequence: 1 MTKREKPFGLRSLFPIH.....AQSLKQEVNQEDSAPASS 515

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2144	81.5	528	2 Q96EV3	Q96EV3 chlamydia t
2	1358.5	51.7	644	10 Q9AS91	Q9AS91 oryza sativ
3	1302.5	49.5	671	8 Q8WH67	Q8WH67 galidieria s
4	1302	49.5	588	8 Q8MFR8	Q8MFR8 citrullus hydr
5	1076	40.9	498	16 Q92J16	Q92J16 rickettsia
6	1020	38.8	514	2 Q8KPN2	Q8KPN2 holospora o
7	976.5	37.1	501	16 Q92HP9	Q92HP9 rickettsia
8	960.5	36.5	540	2 Q9X9D2	Q9X9D2 chlamydia t
9	907.5	34.5	499	16 Q92G15	Q92G15 rickettsia
10	907.5	34.5	511	16 Q92HV4	Q92HV4 rickettsia
11	827	31.4	507	16 Q92J98	Q92J98 rickettsia
12	366	13.9	543	5 Q8SUG0	Q8SUG0 encephalito
13	351	13.3	553	5 Q8SUF9	Q8SUF9 encephalito
14	338.5	12.9	559	5 Q8SRA2	Q8SRA2 encephalito
15	325	12.4	536	5 Q8SUG7	Q8SUG7 encephalito
16	289.5	11.0	105	10 Q945F3	Q945F3 medicago sa

17	229.5	8.7	441	16 Q9PCP0	Q9PCP0 xylella fas
18	203.5	7.7	454	16 Q8P6C3	Q8P6C3 xanthomonas
19	197	7.5	496	10 Q9FLC3	Q9FLC3 arabidopsis
20	188.5	7.2	453	16 Q8P142	Q8P142 xanthomonas
21	173.5	6.6	928	16 Q84237	Q84237 chlamydia t
22	162	6.2	538	16 Q8NU46	Q8NU46 corynebacte
23	161	6.1	918	16 Q9PKG0	Q9PKG0 chlamydia m
24	158.5	6.0	925	16 Q928P5	Q928P5 chlamydia p
25	157.5	6.0	925	16 Q9J516	Q9J516 chlamydia p
26	151	5.7	1002	16 Q8YNE4	Q8YNE4 anabaena sp
27	147	5.6	533	17 Q8TW21	Q8TW21 methanosaic
28	145	5.5	540	16 Q9KSH7	Q9KSH7 vibrio chol
29	144	5.5	660	16 Q9JUS7	Q9JUS7 neisseria m
30	144	5.5	675	16 Q9JZ61	Q9JZ61 neisseria m
31	141	5.4	563	16 Q8PUB5	Q8PUB5 corynebacte
32	140	5.3	474	16 Q9KFD0	Q9KFD0 bacillus ha
33	140	5.3	475	2 Q929P3	Q929P3 bacillus ha
34	137	5.2	577	8 Q94Y38	Q94Y38 cryptocodi
35	137	5.2	498	2 Q33750	Q33750 synechococc
36	136.5	5.2	465	16 Q8FHZ0	Q8FHZ0 escherichia
37	135.5	5.2	441	16 Q916P7	Q916P7 pseudomonas
38	135	5.1	404	16 Q8G4F6	Q8G4F6 bifidobacte
39	134.5	5.1	976	17 Q9V250	Q9V250 pyrococcus
40	134	5.1	516	16 Q8FPUL	Q8FPUL corynebacte
41	133.5	5.1	526	2 Q8MID5	Q8MID5 chaetospae
42	132.5	5.0	552	2 Q9APX1	Q9APX1 pseudomonas
43	132	5.0	463	16 Q8XDD0	Q8XDD0 escherichia
44	131.5	5.0	567	17 Q97VC8	Q97VC8 sulfolobus
45	131.5	5.0	434	16 Q8UDN2	Q8UDN2 agrobacteri
46	131	5.0	420	2 Q46305	Q46305 clostridium
47	130.5	5.0	976	17 Q74088	Q74088 pyrococcus
48	129.5	4.9	401	16 Q8BLQ1	Q8BLQ1 oceanobacill
49	128.5	4.9	446	16 Q8BZT6	Q8BZT6 yersinia pe
50	129.5	4.9	581	16 Q98Q22	Q98Q22 mycoplasma
51	129.5	4.9	484	8 Q47458	Q47458 cryptocodi
52	129	4.9	721	10 Q9FRN4	Q9FRN4 oryza sativ
53	129	4.9	435	16 Q99WS1	Q99WS1 staphylococ
54	128.5	4.9	665	16 Q9ATF1	Q9ATF1 caulobacter
55	128.5	4.9	462	17 Q97XMO	Q97XMO sulfolobus
56	128	4.8	496	5 Q19676	Q19676 caenorhabdi
57	127.5	4.8	397	2 Q84999	Q84999 streptococc
58	127	4.8	462	16 Q8XJY3	Q8XJY3 escherichia
59	127	4.8	999	2 Q93JY2	Q93JY2 errinia chr
60	126.5	4.8	397	2 Q9R925	Q9R925 streptococ
61	126	4.8	513	10 Q94EC4	Q94EC4 oryza sativ
62	126	4.8	548	3 Q9P3K6	Q9P3K6 neurospora
63	126	4.8	464	16 Q8DWD1	Q8DWD1 streptococ
64	125.5	4.8	465	16 Q8ZP36	Q8ZP36 salmonella
65	125.5	4.8	462	16 Q8PHU0	Q8PHU0 escherichia
66	125	4.8	587	11 Q35055	Q35055 rattus norv
67	125	4.8	685	16 Q8EJ17	Q8EJ17 shewanella
68	125	4.8	1185	16 Q8EGS8	Q8EGS8 shewanella
69	125	4.7	391	16 Q9ZCM6	Q9ZCM6 rickettsia
70	124	4.7	501	8 Q94PS6	Q94PS6 cryptocodi
71	124	4.7	518	17 Q8THB8	Q8THB8 methanosaic
72	124	4.7	729	16 Q8F0Z0	Q8F0Z0 leptospira
73	124	4.7	445	16 Q8NVD5	Q8NVD5 staphylococ
74	123.5	4.7	432	16 Q8RGF7	Q8RGF7 fusobacteri
75	123.5	4.7	525	8 Q9TCAG	Q9TCAG nephroselm
76	123.5	4.7	531	8 Q21293	Q21293 reclinomona
77	123.5	4.7	466	17 Q8TUI1	Q8TUI1 methanosaic
78	123	4.7	393	2 Q8KWT2	Q8KWT2 bacillus su
79	122.5	4.7	402	16 Q8P7D9	Q8P7D9 xanthomonas
80	122.5	4.7	465	16 Q8Z7G0	Q8Z7G0 salmonella
81	122.5	4.7	482	10 Q9SWR7	Q9SWR7 cylindrothe
82	122.5	4.7	512	16 Q918V7	Q918V7 pseudomonas
83	122.5	4.7	713	2 Q918D2	Q918D2 polyangium
84	122.5	4.7	1139	16 Q8ZC91	Q8ZC91 yersinia pe
85	122.5	4.6	643	17 Q29273	Q29273 archaeoglob
86	122	4.6	482	10 Q9SWR6	Q9SWR6 cylindrothe
87	121.5	4.6	502	16 Q9CMU0	Q9CMU0 pasteurilla
88	121.5	4.6	534	2 Q9R6U5	Q9R6U5 synechococc
89	121.5	4.6			

90 121.5 4.6 539 2 Q59631
 91 121 4.6 425 16 Q8RCV2
 92 121 4.6 523 16 Q8DAG1
 93 121 4.6 527 8 Q47563
 94 120.5 4.6 393 2 Q8KWS7
 95 120.5 4.6 467 8 Q99295
 96 120.5 4.6 472 16 Q990U7
 97 120.5 4.6 472 16 Q8NUG9
 98 120.5 4.6 482 10 Q8SMR8
 99 120.5 4.6 507 16 Q8XNM3
 100 120.5 4.6 512 10 Q94EC3

ALIGNMENTS

RESULT 1

Q9S6V3 PRELIMINARY; PRT; 528 AA.
 AC Q9S6V3
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
 DE Nucleoside triphosphate transport protein 1.
 GN NP1.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxId=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99138740; PubMed=9973346;
 RA Tjaden J, Winkler H.H., Schweppe C., van der Laan M., Moehlmann T.,
 Neuhaus E.;
 RT "Two nucleotide transport proteins in Chlamydia trachomatis: One for
 net nucleoside triphosphate uptake and the other for the transport of
 energy."
 RL J. Bacteriol. 181:1196-1202 (1999).
 DR EMBL, AJ010586; CAB39534.1; -;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR InterPro; IPR000531; TonB_BoxC.
 DR Pfam; PF03219; TBC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 SQ SEQUENCE 528 AA; 58039 MW; 8FBCDD5FC80B0E3 CRC64;

Query Match 81.5%; Score 2144; DB 2; Length 528;
 Best Local Similarity 77.9%; Pred. No. 7.7e-149;
 Matches 409; Conservative 53; Mismatches 49; Indels 14; Gaps 3;

QY 1 MKTEKEKPGKRSFLMPHITHEIKKVLPMFLMFCITFNVTYVLTDTKDTLLVGA PGSGA 60
 DB 1 MTKTEKEKPGKRSFLMPHITHEIKKVLPMFLMFCISFNVTLLDTKDTLLVTPGSGA 60
 QY 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVL 120
 DB 61 EAIPIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVL 120
 QY 121 HPTPADRLQALIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHE 180
 DB 121 HPTPADRLQALIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHE 180
 QY 121 HPTPADRLQALIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHE 180
 DB 121 HPTPADRLQALIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHE 180
 QY 181 AKRFALFGIGANISLASGRAIIVASAKLRASVSGVDPMGISTLLMAMTIVSGLVMA 240
 DB 181 AKRFALFGIGANISLASGRAIIVASAKLRASVSGVDPMGISTLLMAMTIVSGLVMA 240
 QY 241 SYWMINKVNLTPRRYNEPEEMQKKGAKPRKNNKDSFLYLDSPYILLTLVLAIGYC 300
 DB 241 SYWMINKVNLTPRRYNEPEEMQKKGAKPRKNNKDSFLYLDSPYILLTLVLAIGYC 300
 QY 301 INLIIVTWKSLQKLTQYPNNDYSEPMGNFSFTGVSVILIMLVGAGNVIKRGMTLGAIV 360
 DB 301 INLIIVTWKSLQKLTQYPNNDYSEPMGNFSFTGVSVILIMLVGAGNVIKRGMTLGAIV 360
 QY 360 INLIIVTWKSLQKLTQYPNNDYSEPMGNFSFTGVSVILIMLVGAGNVIKRGMTLGAIV 360
 DB 360 INLIIVTWKSLQKLTQYPNNDYSEPMGNFSFTGVSVILIMLVGAGNVIKRGMTLGAIV 360

RESULT 2

Q9AS91 PRELIMINARY; PRT; 644 AA.
 AC Q9AS91
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative plastidic ATP/ADP-transporter.
 GN P0707D10.36 OR P0038D11.15.
 OS Oryza sativa (Rice), and
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacroidae; Oryzae; Oryza.
 OX NCBI_TaxId=4530, 35947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O. sativa; STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0707D10."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O. sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0038D11."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002910; BAB40979.1; -;
 DR EMBL; AP003234; BAC05539.1; -;
 DR Gramene; Q9AS91; -;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TBC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 644 AA; 68960 MW; B3531863E48334AC CRC64;

Query Match 51.7%; Score 1358.5; DB 10; Length 644;
 Best Local Similarity 53.8%; Pred. No. 2.9e-91;
 Matches 275; Conservative 90; Mismatches 135; Indels 11; Gaps 8;

QY 7 KPFGLRSFLMPHITHEIKKVLPMFLMFCITFNVTYVLTDTKDTLLVGA PGSGA 60
 DB 7 KPFGLRSFLMPHITHEIKKVLPMFLMFCITFNVTYVLTDTKDTLLVGA PGSGA 60
 QY 105 QPEKG-KFL-GVEKTKLTKIVPLGLMFCILENTYITLDRDVLVYAKSSAEIIPFL 161
 DB 105 QPEKG-KFL-GVEKTKLTKIVPLGLMFCILENTYITLDRDVLVYAKSSAEIIPFL 161
 QY 67 KFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVLHPTBEA 126
 DB 67 KFWLVVPCAIIFMLIYAKLSNLSKQALFYAVAGTPELFFALFPVITYPLRDVLHPTBEA 126
 QY 162 KTWVNLPMALIGFMLLYTLNLSREALFYVIFPELFFAGFAVLVPLRVNHPALTA 221
 DB 162 KTWVNLPMALIGFMLLYTLNLSREALFYVIFPELFFAGFAVLVPLRVNHPALTA 221
 QY 127 DLQAIIIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHEAKRFYA 186
 DB 127 DLQAIIIPGGLGLVAIIKRWTPAAFYVLAELMGSVMLSMFGFANEITKHEAKRFYA 186
 QY 222 DKLALALPSPFLGPALIRIWSFCIFYMAELMGSVVSVLFWGPNQITTYEAKERYP 281
 DB 222 DKLALALPSPFLGPALIRIWSFCIFYMAELMGSVVSVLFWGPNQITTYEAKERYP 281
 QY 187 LFGIGANISLASGRAIIVASAKLRASVSGVDPMGISTLLMAMTIVSGLVMA SYWMIN 246
 DB 187 LFGIGANISLASGRAIIVASAKLRASVSGVDPMGISTLLMAMTIVSGLVMA SYWMIN 246
 QY 282 LFGIGANIALIFSRTGVYFYSNKRITLGPIDGMEVSLKGMMSLVLLGLVITSYWGYN 341
 DB 282 LFGIGANIALIFSRTGVYFYSNKRITLGPIDGMEVSLKGMMSLVLLGLVITSYWGYN 341
 QY 247 KNLVLTDPFRYNEPEEMQKKGAKPRKNNKDSFLYLDSPYILLTLVLAIGYCINLLEV 306
 DB 247 KNLVLTDPFRYNEPEEMQKKGAKPRKNNKDSFLYLDSPYILLTLVLAIGYCINLLEV 306
 QY 342 KVLINDPBL--PKSPRRKKKD-KPKLGKESIKVLSSRYVRLATLVAVAGISINLVEV 398
 DB 342 KVLINDPBL--PKSPRRKKKD-KPKLGKESIKVLSSRYVRLATLVAVAGISINLVEV 398


```

Qy 307 TWKSOLKL-OYNNNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKFGMLTGAULTVPVPMV 365
Db 399 TWKSOLKLQOFPSPNEISSFMGDSFTATGIAITFTMML-LGRIILKFGMGAUAAITTPVL 457
Qy 366 LITGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYI 425
Db 458 LITGVGFSLILFQGPLTPMLATMGMTPLLAAYVYGAIONIFSKSAKXSLFDPCKEMAYI 517
Qy 426 PLDDEOK-VKGAIDVVAARFGSGGALLIOGGLVITGSGTGMTPYLAIVLFLIAIWL 484
Db 518 PLDEDMKVVKGAIDVVCNPLGKSGGALLIOGFMILTFGSLANSTPYLGILLYIVLAWL 577
Qy 485 VSATKLNLF--LAQSALKEQVAOEDSAPA 513
Db 578 GAASSLDDQFSSSLAKEDLKRDMSAKKVDPS 608

RESULT 3
Q8MH67 PRELIMINARY; PRT; 671 AA.
AC Q8MH67;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Plasticidic ATP/ADP transporter (Fragment).
GN AATPI.
OS Gaidieria sulphuraria.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Gaidieria.
OX NCBI_TaxID=130081;
RN [1]
RP SEQUENCE FROM N.A.
RA Stemme C., Neuhaus E.H.;
RT "Analysis of the plasticidic ATP/ADP transporter from the red algae
RT Gaidieria sulphuraria."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ251356; CAC80882.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TLC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
KM Chloroplast.
FT NON TER
SQ SEQUENCE 671 AA; 73829 MW; 2025A868F2CFFFD7 CRC64;

Query Match 49.5%; Score 1302.5; DB 8; Length 671;
Best Local Similarity 53.2%; Pred. No. 3.7e-87;
Matches 259; Conservative 84; Mismatches 135; Indels 9; Gaps 5;

Qy 22 HELKKVLPMLMFCITFNNTVYVLDTKDTLIVGAPSGAEAIPIKFMVLVPCAIIFMLI 81
Db 171 NELKKIIPLGVMFMIIFNTYILRDTKDLVAVTT--TGAEIIPFKYANLPGAVLFTIA 228
Qy 82 YAKSLNISKOALFAVGTPLIPALPPYIVYPLRDVLHPTFADRLQALTPGLGLV 141
Db 229 YSKLSNNEDEKTLTYVCTVPIPLFLFSAFVLYPLRHALHPYAFDMVAGVLPASFFAPL 288
Qy 142 AILBNMTFAAFYVLAELMGVYMLSLMFWGFANEITKIHAEKRYALFGIGANISILASGR 201
Db 289 GIINMTFALFYTLAEIMGSVYVSLFWGFANEVTSVDEAKKYPLPGLVANVALIFSGQ 348
Qy 202 AIWVASKLRASVSEGVDPWGISLRILAMNTVSGILVMASTYWNINKVLTTPRFRNPEEM 261
Db 349 YVRFVSDIRRLPGVDGWSGLKYLMTMIGVAGSVIILAMLYIQKNVLTDP--NCVDM 405
Qy 262 QKGGK-GAKPYNNMKSDFLYDRSPYILLLTLVAVGICINLI-EVTKSOLKLQYPMN 320
Db 406 SKRKTKTKTKGLKESAVYLAKSAYINLALVLAIVAGSINIVAEVSKSKLKEAFDPEN 465
Qy 321 DYSEFMGNFSFWTGVSVLIMLFGVGNVIRKFGMLTGAULTVPVWVLITGIVFPAVIFRN 380
Db 466 SYTFMGWFSCTGTSVT-LIMMLIGRFIFRKFGMGFALITPTVLTGIGTFECULFSR 524

```

```

Qy 361 QASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMAYIPLDDEOKVKGKAID 440
Db 525 QLHFVVGMLTTPMLAVVGAIONILSKSKSLFDPCKEMAYIPLDDEOKVKGKAID 584
Qy 441 VVAARFGSGGALLIOGGLVITGSGTGMTPYLAIVLFLIAIWLVSATKLNLF-LAQSAL 500
Db 585 VIGNPMKSGSGSFLIOGGLIFAVGSLASTPYLAIVYGMVIAAARSLDKQF--QEM 642
Qy 501 KEQEVAAQ 507
Db 643 KRSEVAAE 649

RESULT 4
Q8MEF8 PRELIMINARY; PRT; 588 AA.
AC Q8MEF8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Plasticidic ATP/ADP transporter (Fragment).
GN Citrus hybrid cultivar.
OS Citrus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=171250;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Weiss D., Goldschmidt E.B.;
RT "Citrus mRNA for plasticidic ATP/ADP transporter."
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY098893; AAM29152.1;
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TLC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
KM Chloroplast.
FT NON TER
SQ SEQUENCE 588 AA; 64735 MW; B4B6575BB36B37B CRC64;

Query Match 49.5%; Score 1302; DB 8; Length 588;
Best Local Similarity 51.1%; Pred. No. 3.5e-87;
Matches 261; Conservative 87; Mismatches 141; Indels 22; Gaps 5;

Qy 6 EKP--FGLKRSFLMPIHTHELKKVLPMLMFCITFNNTVYVLDTKDTLIVGAPSGAEAI 63
Db 72 EKPKEFG-----IEVTFPKIIPGLMFCITFNNTYILRDTKDLVAVTAKSSAEII 123
Qy 64 PRIFKMLVPCAIIFMLIYAKSLNISKOALFAVGTPLIPALPPYIVYPLRDVLHPT 123
Db 124 PFLTWNVLPALIGFMILLYTKLADKLSKOALFYVYVIFFAFFAFGLYPLLSNALHPQ 183
Qy 124 EFADRLQALIPGLGLVAILBNMTFAAFYVLAELMGVYMLSLMFWGFANEITKIHAEK 183
Db 184 ALADKLNIICPRPLGLPALIRIHSFCLFYMAELMGSVYIVFWGFANQITTVDEAKR 243
Qy 184 FYALFGIGANISILASGRAIVWASKLRASVSEGVDPWGISLRILAMNTVSGILVMASTY 243
Db 244 FYPLFGGANVALIFSGRTVYKFSNLRKNIGPVDGAVSLKGMWSIVLMGLSIGLYW 303
Qy 244 WINKVLTDRFRNPEEMQKGAARKNMKSDFLYDSPLYLLTLVIAVAGICINL 303
Db 304 WNNNVELPFR-----SKKKKEKMGTMESLKEPLVSSRYRDLATVVAVGISINL 355
Qy 304 IEVTKSOLKLQYNNNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKFGMLTGAULTVPV 363
Db 356 VEIVTKSOLKLQOFPSPNEISSFMGDSFTATGIAITFTMML-LSQITPKFGVAAKITPT 414
Qy 364 WVLITGIVFPAVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKXALPDSTKEMA 423
Db 415 WVLITGVGFSLILFQGPLPALAKFGMTPLLAAYVYGAIONIFSKSAKXSLFDPCKEMA 474

```

QY 424 YIPDQKVKGAIDVVAARFGSGGALLIOGGLVIGSGIGAMTPYLAIVLFIITIM 483
 DB 475 YIPDQKVKGAIDVVAARFGSGGALLIOGGLVIGSGIGAMTPYLAIVLFIITIM 534
 QY 484 LVSATKLNKFLAOSGALKEQVADSDAPAS 514
 DB 535 LGAARSLDTQF---TALROEELKEKEMERAA 562

RESULT 5

Q92J16 PRELIMINARY; PRT; 498 AA.
 ID 092J16; 092J16;
 AC 092J16;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE ADP, ATP carrier protein.
 GN Tlci OR RC0081.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 CX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Remesto-Audifiren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 DR EMBL: AE008575; AAL02619.1;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; Tlc; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ COMPLETE PROTEOME.
 KW SEQUENCE 498 AA; 56661 MW; 975BEB9C5BCC56EB CRC64;

Query Match 40.9%; Score 1076; DB 16; Length 498;
 Best Local Similarity 43.6%; Pred. No. 1e-70; Indels 20; Gaps 7;

Matches 213; Conservative 103; Mismatches 153; Indels 20; Gaps 7;

QY 11 KLRSLFWLHITHELKVLPMFLMFCITFNVTYVLRDVTDLVIGAPSGAEALPIPIKWL 70
 DB 12 ELSKIWIIEYENKFLPMAFMFCILNYSITLSIDGFPV--TDGAEALISLKYI 69
 QY 71 VVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFALFPYIVYLPDLVLPTEFADRLQ 130
 DB 70 VLPQVIMVIVYKLCIDILKQENVYVITSFPLGFALFAFVLYPDLVHPDPTISW 129
 QY 131 AILPGLGLVAILRNWTFPAFYVLAELMGVMSLIMWGPANETTKTHEAKRFALGCI 190
 DB 130 SVAAPNVWFIIRIVKWSFASFTYMAELMGVMSLIMWGPANQTKTDEAKRFYSMGL 189
 QY 191 GANISILASGRAIYVASKLRASVSEGVDPWGISIRL---LMAMTIVGLVIMASYMNI 246
 DB 190 LANLALPTYSITIGCLHEKQI-----VAHLKRVPLFVIMTISPLVIILTRYMN 241
 QY 247 KNVLTDPFYPNEEMQKKGAKPRNNKDSFLYIDRSPYILLTLVIAVIGICINLIEV 306
 DB 242 KNVLTDPFYPNEEMQKKGAKPRNNKDSFLYIDRSPYILLTLVIAVIGICINLIEV 306
 QY 307 TWKSGQLQYENMNDYSEFMGNFSFWGVSVLIMLFGGNVIRKFGMLTGALVTPNVVL 366
 DB 300 VMSKSVKELYPTKEAYTITMGEFQYQGVNVAIAFWL-IGSNILKRVSWLTAAMTIPML 358
 QY 367 LTGIIVFALVIFRNO-ASGLVAMFGTTPMLAVVGAIONISLSTKTYALPDSTKEMAYI 425
 DB 359 ITGAAPFAFIFFDSVIAHNLGILASGPLALAVMGIONVYSKVKVSLSDATGMAYI 418
 QY 426 PLDQKVKGAIDVVAARFGSGGALLIOGGLVIGSGIGAM--TPYLAIVLFIITIM 483
 DB 419 PLDQKVKGAIDVVAARFGSGGALLIOGGLVIGSGIGAM--TPYLAIVLFIITIM 478

QY 484 LVSATKLNK 492
 DB 479 IYAVKGLNK 487

RESULT 6

Q8KPN2 PRELIMINARY; PRT; 514 AA.
 ID Q8KPN2; 08KPN2;
 AC Q8KPN2;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Non-mitochondrial nucleotide transport protein.
 GN NMT.
 OS Holospora obtusa.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Holosporaceae; Holospora.
 CX NCBI_TaxID=49893;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Linka N., Hurka H., Lang F.B., Burger G., Winkler H.H., Stamme C.,
 RA Urbany C., Seil I., Kusch J., Neuhaus E.H.;
 RT "Phylogenetic relationships of non-mitochondrial nucleotide transport
 proteins in bacteria and eukaryotes";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY120885; AAM0566.1;
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; Tlc; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 514 AA; 57531 MW; BEC3E24B9289E82F CRC64;

Query Match 38.8%; Score 1020; DB 2; Length 514;
 Best Local Similarity 39.0%; Pred. No. 1.3e-66;

Matches 206; Conservative 118; Mismatches 168; Indels 36; Gaps 12;

QY 1 MTKEERPGKLRSGFLPIHTELKVLPMFLMFCITFNVTYVLRDVTDLVIGAPSGA 60
 DB 1 MSNLSRMSKWRRLWVPVPEMSIFLPMVINFSLFNTVARNVDSLVNPSGSG 60
 QY 61 EAIPIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFALFPYIVYLPDLV 120
 DB 61 KVLPMVNLGVLTPCSISFAVIGYAKLSNLSKQALFYAVGTPFLIFALFPYIVYLPDLV 120
 QY 121 H-PHEPDRLOAILPGLIGVAILRNWTFPAFYVLAELMGVMSLIMWGPANETTK 179
 DB 121 NLSVDWIKKCSQIYPL-LKDFPRAVAYNYSLEFYMALMGVMSLIMWGPANETTK 179
 QY 180 EAKRFVALFGIANISLASGRAIYVASKL--RASVSEGVDPWGISIRLMAMTIVGLV 237
 DB 180 QAKRFYVYVGSWNLGLVAAAGSLTAFSENFILKPERLASGEKDPQRLRYCSCNVVACIA 239
 QY 238 LMASYMINKNVLTDPFYPNEEMQK--KKG--AKPRNNKDSFLYIDRSPYILLTL 293
 DB 240 IVVSYVMNLAKVLTD-----AQKGTDAKGRSKPRLTKESAVYLLKSRMYIAIL 291
 QY 294 VIAIGICINLIEVWKSQKQYR-----NMNDYSEFMGNFSFWGVSVLIMLFGG 346
 DB 292 MLAGGICINLIEVWKSQKQYR-----NMNDYSEFMGNFSFWGVSVLIMLFGG 346
 QY 347 NVIRKFGMLTGALVTPVNVLLTGIVFALVIFRNOASGLVAMFGTTPMLAVVGAIONI 406
 DB 351 NFVOSLSGRPSANITPVTMTIGGFILFVFKOYLGGLGCLFSTSLTSLVAVGLGVNV 410
 QY 407 LSKSTYALPDSTKEMAYIPLDQKVKGAIDVVAARFGSGGALLIOGGLVIGC--- 462
 DB 411 LTKGVKXALFPTKEMAYIPLDQKVKGAIDVVAARFGSGGALLIOGGLVIGC--- 466
 QY 463 GSIG---AMTPYLAIVLFIITIMVLSATKLNKFLAOSGALKEQVADSDAPAS 507
 DB 467 GGLIGIPOLSVPLLGGGLVIGSGIGAMTPYLAIVLFIITIMVLSATKLNKFLAOSGALKEQVADSDAPAS 513

RESULT 7
092HP9 PRELIMINARY; PRT; 501 AA.
AC 092HP9
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ADP carrier protein.
GN Trc3 OR R00722.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
RL Science 293:2093-2098(2001).
DR EMBL; AE008630; AL03260.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
DR TrEMBL; TIGR00769; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 501 AA; 57104 MW; 2EC599FAD865EA3B CRC64;

Query Match 37.1%; Score 976.5; DB 16; Length 501;
Best Local Similarity 40.0%; Pred. No. 1.9e-63;
Matches 199; Conservative 106; Mismatches 174; Indels 19; Gaps 8;

QY 9 FGLKRSFLMPHITHLKLKVLPMFLMFCITFTNTVLTDRDTKTLIVGAPSGAEAIPIKF 68
DB 8 FEKXKEIMWPIERKELKFIPIWALMLCLIFNFGLRSIKDSLIV--PSMGAEIISFLKL 65
QY WLVPVCAITFMLIYAKLSNLSKQALFYAVGTPFLIFALPFTVYLPDVLHPT-EPAD 127
DB 66 WLVPDSCVITFVLVYKLSNLSNMFETIYIIVSGFLPFLPFIYIIPMODIHPDEMINT 125
QY 128 RLQALPPLGLGLVALILRNMTFAFYVLAELMGVYLSLMFGFANEITKIHAKRFYAL 187
DB 126 KLIASY-ENFKMFITIGSOWSALMYIFAEILMSAVINLMFGQFANHLPDTSKARFPV 184
QY 188 FGIGANISLLASGRAYWASKLRASVSGVD-----WGISRLIMAMTVSGVLV 239
DB 185 LGMVGNIGLIIAGSVLVFFSSGQDIVIDSELPDSFNSAGNAIMIQPMSIIVTAGIIM 244
QY 240 ASYMMINKNVLTDPFVNEEMQKKGAKPRMNMKDSFVLIDRSPYLLLTLLVIAIGI 299
DB 245 LLEFRIINFIILDS--INVLDKAKVTAAOKTKLVSIESIKLVHSKYGRIALLLICGL 302
QY 300 CINIIEVTWKSQKLQYENNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGWLTGAL 359
DB 303 LINVIEGWKAKIKELHNTIDYVNFGRFINIMGISCVTMI-IGSNILRLGLISAL 361
QY 360 VTPVWVLTGIVFPAVIFRNOASGLVAMFGTTPMLAVVGAIONISKSTKALPDST 419
DB 362 LTPIMLSITGMFMFIIFIEIGECFDPNL--LYAAIIVGAIIONISKSKSYLFDST 419
QY 420 KEMAYIPIIDOKYKGAIDVVAARPGKSGALIQGLLVY--CGSIGAMTPYAVLL 477
DB 420 KEMAYIPIIDOKYKGAIDVVAARPGKSGALIQGLLVY--CGSIGAMTPYAVLL 479
QY 478 FIATWLVSATKLNKLF 495
DB 480 VMMSLMINVIKLNKEY 497

RESULT 8
09X9D2 PRELIMINARY; PRT; 540 AA.
ID 09X9D2

AC 09X9D2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Nucleoside triphosphate transport protein 2.
GN NPT2.
OS Chlamydia trachomatis.
OC Chlamydia; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99138740; PubMed=9973346;
RA Tjaden J., Winkler H.H., Schweeppe C., van der Laan M., Moehlmann T.,
RA Neuhaus E.;
RT "Two nucleotide transport proteins in Chlamydia trachomatis : One for
RT net nucleoside triphosphate uptake and the other for the transport of
RT energy";
RL J. Bacteriol. 181:1196-1202(1999).
DR EMBL; AJ010587; CAB39535.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; T1C; 1.
DR TrEMBL; TIGR00769; AAA; 1.
SQ SEQUENCE 540 AA; 59648 MW; C43C2A3FC4226FD CRC64;

Query Match 36.5%; Score 960.5; DB 2; Length 540;
Best Local Similarity 37.5%; Pred. No. 3.1e-62;
Matches 198; Conservative 115; Mismatches 194; Indels 21; Gaps 6;

QY 4 TEKRPGLKRSFLMPHITHLKLKVLPMFLMFCITFTNTVLTDRDTKTLIVGAPSGAEAI 63
DB 3 SEVSPSFRQYPPPIRSEFSKPIPLFLAFVGVNVALKTKRDSLIVGSRGAEVI 62
QY 64 PFIFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFALPFTVYLPDVLHPT 123
DB 63 PFLVWGVIPGAVIVTMIYGMMSRRYSGVIFISLVGGLFPLFAVIVIPIDGALHN 122
QY 124 EFADRLQALPPLGLGLVALILRNMTFAFYVLAELMGVYLSLMFGFANEITKIHAKR 183
DB 123 KLAALQSLPDPGGRGVWQVMSYSLIYVMSLSVSLTFLWGVANHITSVREAGR 182
QY 184 FYALFGIGANISLLASGRAYWASKLRASVSGVD-----WGISRLIMAMTVSGVLV 242
DB 183 FYALINGLMSSVFAVSVLSMLGRSPVIAFPAVDPMHEMLNITLLIIVLAGGVITLV 242
QY 243 TWINKNVLTDPFVNEEMQKKGAKPRMNMKDSFVLIDRSPYLLLTLL 293
DB 243 QKLR--LMDETSMLEGLAEMSVADLKKKKRSKAK--AKSIFALLIRSRYLIGIAV 298
QY 294 VIAIGICINLEVTWKSQKLQYENNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFG 353
DB 299 VLSYNLVHLEFVWVKQVCRIVASRVENSYMSRITTLTGIVBALGIFPAAGOTIRWG 358
QY 354 WLGTALVTPVWVLTGIVF-ALVIFRNOASGLVAMFGTTPMLAVVGAIONISKSTK 412
DB 359 MTGVALVPLTILITLIGALFEGAIYAVGDMIFGGIIGISPLVITAMLGQVQNFSAIK 416
QY 413 YALPDSIKENAYITLDOQKVKGAIDVVAARPGKSGALIQGLLVYCGSIGAMPYL 472
DB 419 FTVPDQTEKMAFIPLDEDEKRYGKAIDGVISRVKSGGSLVYQGLIIFSSVAASINAI 478
QY 473 AVILLFIATWLVSATKLNKLF 495
DB 479 TIVLLLAGSWIFVIAMVGRBYTAKTETLVVNASBEDVLOERBAS 526

RESULT 9
092G15 PRELIMINARY; PRT; 499 AA.
ID 092G15
AC 092G15
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ADP carrier protein.

GN Tlcs OR RC1138.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 DR EMBL; AE008663; AA03676.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; Tlcs; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Complete proteome.
 SQ SEQUENCE 499 AA; 56918 MW; CD873E469763E2AF CRC64;
 Query Match 34.5%; Score 907.5; DB 16; Length 499;
 Best Local Similarity 37.4%; Pred. No. 2.2e-58;
 Matches 190; Conservative 107; Mismatches 176; Indels 35; Gaps 13;
 QY 1 MKTKEKRP-GKLRSEFLPIHTEHLKKVLPMFLMPCITENYTVLRDTKDLIVGAPSG 59
 DB 1 MISTSRSRKRFRAFPVHVEYELGKFTPMSTLFCILFNQVNRILKDSISE--IS 58
 QY 60 AEAIFIKFWLVVPCALIFMLIYAKLSNLSKQALFYAVGTPELIFPALFPYIYPLRDV 119
 DB 59 AEAIFAKYCVTPAALFVLIYAKMINVLFTEKIFVYLAFIFSPVLFYVYIPNHI 118
 QY 120 --LHTEFEDRLQALLPGLGLVALNRWTPAAYVLAELMGVSLMFGFANEITK 177
 DB 119 FVHHPNNLADWERY--PHFKYISLVGMGYIVYSIAELWPNIFVYLLFWQFANEILT 176
 QY 178 IHEARFPAVPGIGANISILASGRAIVAS-----KLRAVSEG-VDPWGISLMLAM 230
 DB 177 TEEARFYTFLSPGNSLLIVGFLMMNLISSEDTIIFKFMSTISDKITLVQSTTVIAIV 236
 QY 231 TIVSGLVMAASYWYMNKNTLDRPFYNPEMOKGKG--AKPKNMKDSFLYLDSPYL 288
 DB 237 AITCLLVK---FISKVFTNPLFY-----AKASGRSTSERMGLISFKYIAKSKLM 287
 QY 289 LITLIVAYGICINILEVTWKSQKLQYPPNADYSEPMGNESFPTGVVSVLIMFVGAV 348
 DB 288 LLLISAAFGPAINLVEAVWKAKIKELYPTVTVAEFNSLYILMTG-VAIMVMTIIGNNI 346
 QY 349 IRKFGMLGALVTPVWVLTGIVFPALVIFRQA-----SGLVAMFGTTPMLMAVVGAIQ 404
 DB 347 MRMHNPFAVAIVSPIYIMWTGILFVLIVFDQIISLPGAILM--SPALAVSIGIQ 403
 QY 405 NILSKSTKALFDSTKEMAYIPLDOEKVKGKRAIDVVAARFGKSGALLIQGL--LVIC 462
 DB 404 NILAKTKYSIMWTSREMLYIPLDEELKTKGKAADVISAKEYSSGSLVGSIIFTLVPT 463
 QY 463 GSIGMTPLVAILLFIATIMLVSAATKL 490
 DB 464 ATFTLISPLMVVFTVCLAMTVAVRKI 491
 RESULT 10
 Q92HV4 PRELIMINARY; PRT; 511 AA.
 AC Q92HV4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE ADP_ATP carrier protein.
 GN Tlcs OR RC0066.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 DR EMBL; AE008626; AA03204.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; Tlcs; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Complete proteome.
 SQ SEQUENCE 511 AA; 58036 MW; E194D7E21591EB02 CRC64;
 Query Match 34.5%; Score 907.5; DB 16; Length 511;
 Best Local Similarity 37.9%; Pred. No. 2.2e-58;
 Matches 187; Conservative 111; Mismatches 175; Indels 21; Gaps 9;
 QY 9 FGKRSFLMPPIHTEHLKKVLPMFLMPCITENYTVLRDTKDLIVGAPGAAIPIKF 68
 DB 19 FSKLTIDWPIKRHSVSKFLFTILMFCILFIQNLIRALKDSIYTM--IGAEIISFKF 76
 QY 69 WLVPICALIFMLIYAKLSNLSKQALFYAVGTPELIFPALFPYIYPLRDVH--PTFEA 126
 DB 77 WGVMSAPLTAIYKLVKVRKKAENIFYLIIISFLTFPFAAYVIFPHENLHSPVTQ 136
 QY 127 DRLQALIPGLGLVALIRNTPFAFYVLAELMGVSLMFGFANEITKIHAKRFA 186
 DB 137 NLMAST--PNLKWPIWLSKSPSLFYIABLNNVYFALLFWQFVNNITVESKRYP 194
 QY 187 LFGIGANISILASGRAIVASKLRASVSEGVDPNG---ISRLIMANTIVSGLVASY 242
 DB 195 LFGILSGTGIYLAQCFENLSINDYVNRKFKALOSSFHTLSIQIITLVILIGIATKF 254
 QY 243 WINKNVITDRPFYNPEMOKG--KGAKPKNMKDSFLYLDSPYLITLIVAYGICI 301
 DB 255 WLNHRKVL-----DKEMALLRFKAKKSKWTIASEFQMLSSHRILATLLCYGIAI 308
 QY 302 NLEVTWKSQKLQYPPNADYSEPMGNESFPTGVVSVLIMFVGAVIRKFGMLTGAIVT 361
 DB 309 NLVSGPKAAATKTKYKPTFAAIFGSLYSTGVFTILFV--LGSNIYRRLGWTPAAVIT 367
 QY 362 PNVVLTGIVFALVIFRQASGLVAMF-GTTPMLAVVGAIONLSKSTKVALFDSTK 420
 DB 368 PLIVFTIGILFFAVNNEERFAGLIIANFILDPALIAITIGAIQNVLSKSKYLPDSTK 427
 QY 421 EMAYIPLDOEKVKGKRAIDVVAARFGKSGALLIQGLVIVCGSIGMTPLVAILLEFI 480
 DB 428 EMAYVPLDPETIKIGKKAADVIGTKLGSAGSAPLOSIVFTILPBAVSQISITCMLITFI 487
 QY 481 A--IMLVSAATKLK 492
 DB 488 TCLTMWATKALNK 501
 RESULT 11
 Q92198 PRELIMINARY; PRT; 507 AA.
 AC Q92198;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE ADP_ATP carrier protein.
 GN Tlcs OR RC0522.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=781;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Rensse-Audiffren P., Fournier P.-E., Barbe V.,
 SA Sautou D., Roux V., Coesart P., Weissenbach J., Claverie J.-M.,
 RA Raulot D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 DR EMBL; AB008614; AL03060.1; -
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Complete proteome.
 SQ SEQUENCE 507 AA; 58365 MW; 49A92CD91E300C70 CRC64;

Query Match 31.4%; Score 827; DB 16; Length 507;
 Best Local Similarity 35.8%; Pred. No. 1.7e-52;
 Matches 179; Conservative 106; Mismatches 183; Indels 32; Gaps 9;

QY 11 KLRSEFLMFIHTEHKKVLPFMEFECITENTVTRDTDLVGAPSGAIAPIKFWL 70
 DB 18 KFRITVPIRISYELTKFIPMALMFILLNOLVRSIKDSFVVLIS--EVLSPKLMG 75
 QY 71 VVPCAIIFMLIYAKLSNLSKQALFYAVGTEPLLFFALFPTVIYPLRDVHPTEFADRLQ 130
 DB 76 EMPGILFVIFYSKLONMTTEQVFRITGTFFFAIFGFLFPYRFFHDELIKH 135
 QY 131 AILPGLGLVAILRNTPFAFVYLAELGSMLSMFPGANDITIKHAKRYALFGI 190
 DB 136 ITVPLHLMFLLINGOWSLVIFYINGELMPVIFLLYWLQANKITKEAPRFSFPTL 195
 QY 191 GANISLASGAIYAWASK---LRASVSEGVDPWGISILRMAMTIVSGILMASYWIN 246
 DB 196 FGGNLSLISGVIIYFPAKSEHFLPLSHLNDYHEILKSTIYLLISGLICLALHLKID 255
 QY 247 KNVLTDRFYNPBEEMOKKKGAKP-----KNNKDSFLYIDRSFYILLTLVIAIGICI 301
 DB 256 KSVV-----EADKNIFKQNMIDILKSLVDSAVILTSRYLGICLWMSYSMSI 306
 QY 302 NLIEVTKSQKQKQYPMNDYSEFMGNFSFTGVSVYLIMLFGVGNVIRKGTALVLT 361
 DB 307 SLIEGLMWSKQKQYPAKCPDIAVHGKVFWTGILT--LVSAPLSSILRIKGMFWGAIIT 365
 QY 362 PVWVLTLGVIFPAVIFPNQASGLVAMFG--TTPMLAVVGAIONILSKSTKVALPDSK 420
 DB 366 PIMFGAGWFFSTVEFNHGNIVNTIGSAPLVVIFYGSLHMSKYSLSLFPATK 425
 QY 421 EMAYIPLDQOKVKAIDVVAARFGKSGALLQOGLVLCGSI--GAMTPYALVILF 478
 DB 426 EMVYIPDSEKTKGKAADVDMGAKIGKSGAIIQ---FISFISFPNAINNDIAGILMF 481
 QY 479 ----TIAWVSATKLNKLF 494
 DB 482 SFIIICLLMLYGVKVLKYY 501

RESULT 12
 Q8SUG0 PRELIMINARY; PRT; 543 AA.
 ID Q8SUG0
 AC Q8SUG0
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein ECU10_0520.
 GN ECU10_0520.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_Taxid=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1.
 RX MEDLINE=21576510; PubMed=11719806;
 RA Katicka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
 RA Prensler G., Barbe V., Peyretilade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 Encephalitozoon cuniculi".
 RL Nature 414:450-453(2001).
 DR EMBL; AL590449; CAD25771.1; -
 KM Hypothetical protein.
 SQ SEQUENCE 543 AA; 61647 MW; 44D7AB5F2A55B1A CRC64;

Query Match 13.9%; Score 366; DB 5; Length 543;
 Best Local Similarity 23.7%; Pred. No. 9.2e-19;
 Matches 125; Conservative 119; Mismatches 228; Indels 56; Gaps 15;

QY 5 EEKRFGLRSEFL-----WPIHTEHKKVLPFMEFECITENTVTRDTDLVGAP 56
 DB 23 EEVYKGTGEPFKHVRANRNP-----RVLYSLIFGVITVHTTGMNIREVYLMGR 74
 QY 57 GSGEALPIFIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPALFPTVIYPL 116
 DB 75 -QPMSEFIFKSTFLPFCSLIFMAIOLGSLTPSKMFIITLILFSGCITLFGVWVPL 133
 QY 117 -----RDVLPTEPAD--RLQAILPGLGLVAILRNTPFAFVYLAELGSMVL 164
 DB 134 KGYIQKDFWYSRDI-----FGDGMESLRHLPYVFLVFNEMWTSFLPCEMGMALV 188
 QY 165 SLMPWGANITIKHAKRYALFGANISLASGAIYAWASKLRASVSEGVDPWGISTL 224
 DB 189 SYEFNIFANVSRRQSRQRYIVNLSNALSIFLAVTLVFNKMRGVAFEETKELFRI 248
 QY 225 RLAMNTIVSGILMASYWINKKNVLTDRFYNPBEEMOKKKGAKPNNKDSFLYIDRS 284
 DB 249 LILVLSGTVGILAKKY--MERELPAPVPL--IREVEK--TSBERKLUKDEARQTSRS 304
 QY 285 PYILLTLVIAIGICINLIEVTKSQKQKQYPMNDYSEFMGNFSFTGVSVYLIMLV 344
 DB 305 KLIIAISLNLTLGVSTLVEATFKSGIAGARTNNSKETTFANFYNGLBQIIAISLV 364
 QY 345 GGN-----VIRKSGWLGAIVTPVMVLTIGVFPALVIFNQA--SGLVAMFGI----- 391
 DB 365 VINTPYSALVKKGWKLIAL--DIVAMFSLFVFLAFYVAGADSGANVLFSLFQRM 423
 QY 392 TPLMLAVVGAIONILSKSTKVALPDSKEMAYIPLDQOKVKAIDVVAARFGKSGG 451
 DB 424 PTILENTLGLVTRNASKIKIKYGADVSKAISKQIDPLRYAKYKAVYDGLCGKLSLG 483
 QY 452 ALI---QOGLVLCGSI--GAMTPYALVILFIAIWTVSATKLNKLF 496
 DB 484 SIICVTWGLMDIT--DIRAVSVSGILVITIAMVFFILKYSRQFA 530

RESULT 13
 Q8SUP9 PRELIMINARY; PRT; 553 AA.
 ID Q8SUP9
 AC Q8SUP9
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein ECU10_0540.
 GN ECU10_0540.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_Taxid=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=GB-M1;
 RC MEDLINE=21576510; PubMed=11719806;
 RA Katiinka M.D., Duprat S., Cornilliot E., Metenier G., Thomarat F.,
 RA Prenslér G., Barbe V., Peyretailade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi.";
 RL Nature 414:450-453(2001).
 DR EMBL; AL590449; CAD25773.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 553 AA; 62160 MW; 5091200D1F656BF5 CRC64;

Query Match 13.3%; Score 351; DB 5; Length 553;
 Best Local Similarity 25.0%; Pred. No. 1.3e-17;
 Matches 132; Conservative 89; Mismatches 236; Indels 70; Gaps 17;

QY 14 SFLMP-----ITHELKVLPMFLMFCITENYTVLRDTKDTLIVGAGSGAALPFIKF 68
 DB 29 STWPLSRIRVARCCKWGLGSLAFGASAVIYFSRVMKDSFVLSR--QLPALSFLKT 86
 QY 69 WLVPICALIFMLIYAKL--SNLSKQALEYAVGTFFLIFFALPFTVIYPLRDVLP-----122
 DB 87 CFVLPISIVTVIGVOKLAVTRTISKVPDYTLIASFL--YLLIGNVLLPFAKTIQPGIYF 144
 QY 123 --TEPADRLQAILPPG--LLGLVALRNWTPAAYVLAELMGSVMLSMFGPANEITK 177
 DB 145 SRDIFADDMKAY--KGFEFLFALIFLNMWTSFVYVCAELFGSLVQFMFLAFANALY 202
 QY 178 IHEKRFALFGIGANISLASGRAIYVASKLRASVSGVDPWGISLRLAMT--VSG 235
 DB 203 IROSTRMPLFVYSNILLSSSTFSYK-----KVRWDYKTKCLTNSFEAFVG 255
 QY 236 LVLNASYV--WINKNVLTDPREFYNEEMQKAKPRMMKDSFLYIDRSPLYLLTL 292
 DB 256 AMIATVLYKRYAEETILKQOLFRTBEVAK-KKGRKSAGSBSMKLMAQSKPLVAVM 314
 QY 293 LVIAVGICINLIEVTWKSQKLQYPMNDYSEPMGNSFMTGVVS-----VLIML 342
 DB 315 NALFEYAGTNLIESMKNGISVAADANN--MEKRAVSASIVSGEQRVVALVALILL 369
 QY 343 FVGGNVIRKFGMLNATLTPMVNLTGLVFPALVIFR--NQASGLVAFGTTPLMLAVV 400
 DB 370 TPISITLVQTHGMITMAIVPLVTIVSLVIFGSAFENSNYPEG---KTSVILSSLVK 424
 QY 401 GAIQNIL-----SKTKYALFDPSTKEMAYIPLDQOKVYKKAIDVVAARFGK 448
 DB 425 GYKRFNIECNIGIYCVSGMKIAKAFYDISKEALSLOIDPLRYRLKAVYDGLCGKLGK 484
 QY 449 SGGALLIQGLLVI--CGSIGAMTPYLAVILLFTIALVLSATKLNLF 494
 DB 485 SIGSLYAMFMSVGVNDVRAAPITLGMWLLISPIWISVILNRYK 531

RESULT 14

Q8SRA2 PRELIMINARY; PRT; 559 AA.

AC Q8SRA2;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE ADP/ATP carrier protein 1.
 GN EC008_1300.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryoniadae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA GenomeScope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=GB-M1;
 RC MEDLINE=21576510; PubMed=11719806;
 RA Katiinka M.D., Duprat S., Cornilliot E., Metenier G., Thomarat F.,
 RA Prenslér G., Barbe V., Peyretailade E., Brotier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi.";
 RL Nature 414:450-453(2001).
 DR EMBL; AL590448; CAD26436.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 559 AA; 63516 MW; 9B1558ED28BCD5D CRC64;

Query Match 12.9%; Score 338.5; DB 5; Length 559;
 Best Local Similarity 24.6%; Pred. No. 1.1e-16;
 Matches 134; Conservative 107; Mismatches 237; Indels 67; Gaps 19;

QY 10 GKLSFLMPIHTHEKVLPMFLMFCITENYTVLRDTKDTLIVGAGSGAALPFIKF 69
 DB 31 GILRYF--RVARAEYTFALIGLMLFGIIGFYSFMRILKDMFVN--VRQETTLFIKIF 86
 QY 70 LVVP--CAIIFMLIYAKLSNLSK-----QALFYAVGTPLIFPALPFTVIYPLRD 118
 DB 87 YILFVSMALVFLIQMLGTQVSRIFSLFCGGSFSLFLGAVFLIEHQVSPSK-FLFRD 145
 QY 119 VLAHTEPADRLQAILPPGLGLVALRNWTPAAYVLAELMGSVMLSMFGPANEITKI 178
 DB 146 MFIDGKSSSLNLFKSMFLTNELPLATYF--ISAEMWGSIVLSYLFSLNESCTI 201
 QY 179 HEKRFALFGIGANISLASGRAIYVASKLRASVSGVDPWGISLRLAMTIVSG-LV 237
 DB 202 RQSRFIPPLITINVLFLSATYAGAFPLREKLA-----FQGNVLLSGIIFPGELV 256
 QY 238 LMASYV--WINKNVLTDPREFYNEEMQK--KKGAKPRMMKDSFLYIDRSPLYLLTL 293
 DB 257 VLVIKLYLERVTMKRPLF--VSSGSRKAKANVSFSEGLIEMSGKLLAMSLI 312
 QY 294 VIANIGICINLIEVTWKSQKLQYPMNDYSEPMGNSF--FWGVSVLILVFGGN 347
 DB 313 VLFNINISYNVESTFKGVKAAAEYFN--EEKGKXSGFENRIDQWTSVVAICNLSPSS 371
 QY 348 VIRKFGMLTGLVTPMVNLTGLVFPALVIFR--NQASGLVAFGTTPLMLAVV 400
 DB 372 YVETRGFLVGLIPIYTLMAIVFLGSALINTSMESGLIYVGLFPGGKPLVLENYF 431
 QY 401 GAIQNILSKTKYALFDPSTKEMAYIPLDQOKVYKKAIDVVAARFGKSGALLIQGLLV 460
 DB 432 GVIFMSLLKITKYSAPDICEKLGMRINPTYBARFKSVYDGIFFGLKSGISGI--YGLIM 489
 QY 461 I---CGSIGAMTPYLAVILLFTIALVLSATKLNLF-----LAQSALKEQ 503
 DB 490 FEALDTEDLKRAFTPTGIIIFIFVWVKAIIYLSRSEYSAVQHNRDVIDMTKARKSL 549
 QY 504 EVAQE 508
 DB 550 ETPEE 554

RESULT 15

Q8SUG7 PRELIMINARY; PRT; 536 AA.

AC Q8SUG7;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical protein EC010_0420.
 GN EC010_0420.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryoniadae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.

SQ SEQUENCE 441 AA; 47931 MW; BDD996A58EC4E89 CRC64;
 Query Match 8.7%; Score 229.5; DB 16; Length 441;
 Best Local Similarity 23.1%; Pred. No. 7.7e-09;
 Matches 122; Conservative 84; Mismatches 197; Indels 125; Gaps 27;
 3 KTEEPFKLSFLPIHTELKVLPMELMFCITFNVTYLRDQKDLVIGAPSGSEA 62
 Db 22 RPEEP-----AVLM-----SMLYVALFLAY-----VLRPRDEL--GVAG-GVON 61
 Qy 63 IPIKFMVLVPCALIFMLIVAKLSNLSKO-----ALFYAVGTPFLIPALPFTVIYP 115
 Db 62 LP-----WLFETG-TLLAMLVASPLFALAVRSLPRQGFIALAVRFPAANLVFLAL----- 109
 Qy 116 LRDVLPPEFADRLQALIPGLGLVALLRNWTFAFYVLAELMGSVMSLTMFGFANEI 175
 Db 110 ---LH---FADPOWQV-----WVGRAFIWVSFVFLFVVS--VFMSFWDL 148
 Qy 176 TKHAKRFYALFGANISLLASGRALVWASKLRASSEGVD--PMGISLRLLAMTIV- 233
 Db 149 FDSEQKRLFGIFAAAGT---AGS---LIGSATISGLIEHIDRSW-----LMAIATVF 195
 Qy 234 SGLVMSYWMINKVLTDPREFY-----NPEMOKGKGAAPKKNMKSPLYLDRSPYI 287
 Db 196 LELAVLAS---RLSRITAPAFQHARNDPQPLG-----GIFAGVHTLSRSPYL 243
 Qy 288 LLLTLVLAIGICINLIEVTKSQKLOYPNMNDYSEMGNSFETGVASVLMFVGN 347
 Db 244 GGLAIFILLYSVTSTFLYFOQASIAQASFPDPAATAFANIDLVNATVTLFOLFTVGR 303
 Qy 348 VIRKGMLTGALVTPVWVLTGIVFALVIFRNQASGVAMEGTPMLAVVAGIQNL 407
 Db 304 MAAIVGIVATCVLPVSL-----GPAAL-AASP-SVAIVIAH--OVA 343
 Qy 408 SKSTKYALFDSTKEMAYIPLDQOKVGAKAIDVVAARFGSGALIQGLLVIGSIGA 467
 Db 344 RRVANFALARPARELFTSSAREDRYKAKNFIDTVVRGQDQIASWGVAIGML-GLTLA 402
 Qy 468 MTPILAVILFPIALVWVSATKANKLFLAQSNLKEQVAQEDSAPASS 515
 Db 403 QIPMAIVPL---SAWVIGLSVLTGRTHQAO-----RQDAATAPS 439
 RESULT 18
 ID Q8P6U3 PRELIMINARY; PRT; 454 AA.
 AC Q8P6U3;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Hypothetical protein KCC2872.
 GN KCC2872.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OC NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A. / NCPBP 528;
 RC STRAIN=ATCC 33913 / NCPBP 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 Camarotte G., Canavan F., Cardoso J., Chambeiro F., Clapina L.P.,
 Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Secubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463(2002).
 DR EMBL; AB012400; AAA42144.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 454 AA; 49344 MW; 405A41811E056D7 CRC64;
 Query Match 7.7%; Score 203.5; DB 16; Length 454;
 Best Local Similarity 19.8%; Pred. No. 6.3e-07;
 Matches 101; Conservative 78; Mismatches 152; Indels 179; Gaps 24;
 31 FLMFECTFNVTYLRDQKDLVIG-----PGS-----GAEAFPIKFMVLV--PCALI 77
 Db 30 FLYFCLLSGYVYRPREMGASADVAAIFPAGMIAPFGHGNPLDFTLQVFTCTFL 89
 Qy 78 FML-----IVAKLSNLSKQALF-----YAVGTPFLIPALPFTVIYPLRDVLPTEPADR 128
 Db 90 ITLLQPIYGAIVSRVRRVFLPVYVGFPIAT-LLLFYVFLDTGV----- 133
 Qy 129 LQALIPGLIGVAILRNMTFAFYVLAELMGSVMSLTMFGFANEITKHEARFALF 188
 Db 134 ---FG-RGMAFFLWTGVFNLFV-----AVWMSFVADVFSNAEARSYYGYI 175
 Qy 189 G-----IG-ANISLASG-----RAIV--ASKLRASVE 215
 Db 176 GAAGTGAFLGPIITRTLVERIGIALMLVSAGLGMCCVCLRLMAVARBOEROVTS 235
 Qy 216 GVDPWGISLRLLAMTIVSGLVMSYWMINKVLTDPREFNPEMOKGKGAAPKKNMK 275
 Db 236 GEV-----MGDVLTGKLTIV----- 252
 Qy 276 DSFLYDRSPYLLTLVLAIGICINLIEVTKSQK-klQYPNMNDYSEMGNSFETGV 334
 Db 253 -----REPLRLWAMV--FGVGVGTLLYNEQALVRLTYDDAASTAYANIDLVN 304
 Qy 335 VVSIVLMFVGNVIRKGMVLTGALVTPVWVLTGIVFALVIFRNQASGVAMEGTP 394
 Db 305 ALTVLQTLVLRLLSRFG-----IAPLLIPGAALIFGPVAV--SASPL-----PM 349
 Qy 395 MAAVVAIONILSKSTKYALFDSTKEMAYIPLDQOKVGAKAIDVVAAR----- 445
 Db 350 MAAIV-----QWTRSEPALAKPARETLVTRDREVRVYAGATIDVVVRGGDLVWVI 404
 Qy 446 -----FGKSGGALIQGLLVIGS-STGAM 468
 Db 405 HKVVSAPGSS--VFGVGLVIACGMVIGAL 432
 RESULT 19
 ID Q9FLC3 PRELIMINARY; PRT; 496 AA.
 AC Q9FLC3;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE Genomic DNA, chromosome 5, TAC clone K18123.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Colombia;
 RX MEDLINE=98344145; PubMed=9679202;
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asanizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
 RT features of the regions of 1,381,565 bp covered by twenty one
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:131-145(1998).

DR EMBL: AB010692; BAB09974.1;
SQ SEQUENCE 496 AA; 53606 MW; 7BA868BE7369CDE6 CRC64;

Query Match 7.5%; Score 197; DB 10; Length 496;
Best Local Similarity 22.0%; Pred. No. 2.1e-06;

Matches 119; Conservative 78; Mismatches 203; Indels 140; Gaps 26;

```

OY 19 IHTHKKVLPWFLPFCCTFNTVTLRDTKDLIVGAPSGAEALP--PI-KFWLVPCA 75
DB 17 VHPHLPALHSSSCFFPLSAFVVLPRDE---GALSLSKSLPGLFVGSFLITLRA 73
OY 76 IIFMIIYAKLSNLSKQAL---FYAVGTP--FLIFPALPFTVIYPLRDVLPTEPADR 128
DB 74 PLSTFIFS-LPVLSSKSLVLIHREFSLVLCPILMAA-----SPHSNSK 119
OY 129 LQALIPGGL--GLVAIL---RANTPAAYLALSGVMSLM---FWGPAEITKI 178
DB 120 DAVEASGLTNDGANAATGMDNHGWFYISVRVGFPLV-VALNLVAISSTWARIIDVMDI 178
OY 179 HEAKRFVALFGANISILASGRATVWASKLASVSEGVDPWGISLRLMAMTIVSLVL 238
DB 179 QSGARLPFGVAGATL-----GQLFGSVPAATAM-WGPYLLPFLALMEPPA 225
OY 239 MASYMIN-----KNVLTDFRYNPEEMOKKGAKEPOMMK 275
DB 226 QSSKGTINDISQSSSELSPLRGTDNDHQERKEATSPKVSPP-----KVASFSPIS 278
OY 276 -----DSFLYDRSPYLLITLTV-----IAYGICINLIEVTKSQ--KLQ 315
DB 279 TTRDPFMAILQOMRLILASPYLLLSLFLMGLGAVISSEFFYQKVIITATIKSIGRRRL 338
OY 316 YPNMNDYSEFMGNPSFWTGVSVLIMLFVGNVIRKPFMLGALT--PWVWLLGIYIF 373
DB 339 FAQINSF-----VAVFILTI--GQLT-----LTGRILTVAGVTVAISASPV 377
OY 374 ALVIFRNOASGLVAMFGTTPMLAVVGAIQNLSKSTKVALFDSTKEAAYPLDQEQV 433
DB 378 AL-----GNVLAIAIMPTMTVY---AVSETLAKVTYVTRGRRELFVSGDERY 426
OY 434 KGAALIDVVAARFGSGGALLIOGGL-VICGSGIMPTYLAVILLFIATIMVSTKLNK 492
DB 427 KAKCIDIVIVORLGDAAA-----GLFEVLTALAGQSTASIALPVCLIMVTAFPLGR 482

RESULT 20
O8P142 PRELIMINARY; PRT; 453 AA.
ID O8P142;
AC O8P142;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein XAC3056.
GN XAC3056.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RC MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farh C.S., Furlan L.R.,
RA Queiroz R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavayan F., Cardozo J., Chamberg F., Clapiena L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Curcio-Santos J.R., El-Dorriy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madela A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

```

RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezra R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RT Nature 417:459-463(2002).
DR EMBL: AB011948; AAA37901.1;
KW Hypothetical protein, complete proteome.
SQ SEQUENCE 453 AA; 49247 MW; 34437007F15C2764 CRC64;

Query Match 7.2%; Score 188.5; DB 16; Length 453;
Best Local Similarity 20.6%; Pred. No. 7.9e-06;
Matches 107; Conservative 73; Mismatches 140; Indels 199; Gaps 25;

```

OY 31 FLMPFCITFNTVTLRDTKDLIVGAPSGAALPPIKFWLVPCAIIIMLYAKLSNLS 90
DB 28 FLYEFCILSGYVLRPVRAM--GA-SADVVAL-----FPTGMI----- 63
OY 91 KQALFYAVGTP-----FLIFPALPFTV---IYPLRDVLPTEPADRLQAI 132
DB 64 --AFFAHGVLPKQFTLOVLTCTFLIWLVIQPYGALVSRYPR-----V 107
OY 133 LPPLGLVAILRWTFAPFYVAL-----LWGSV---MLSLMFGPAEITKI 178
DB 108 FLPGYVGFPA-----TLLFVLPDSGVPRGMAFLWVTFNLPAAVAFMSFADVPSN 163
OY 179 HEAKRFVALFG-----IG-ANISLSASG-----RAIVW-- 205
DB 164 AQAASYVYIGAGTLAGFLGPLVLRVLRIGIAHMLVSGFLAVCVCLRLWAV 223
OY 206 ASKLKASVSEGVDPWGISLRLMAMTIVSGVMAVSWYMINKNTLTDRFNPDEMOCK 265
DB 224 AREBGGQSSSEVP-----WGDVVGSKLTV----- 250
OY 266 KGAAPKMMKDSFLYDRSPYLLITLTVIAYGICINLIEVTKSQ--KLQYPMNNDYSE 324
DB 251 -----REPLRLWLAEMVL-FGVGVGLNLNEQALVRLYTDAAAATA 292
OY 325 FMGNPSFWTGVSVLIMLFVGNVIRKFGWLTGALVTPVWVLTGIVFALVIFRNOASG 384
DB 293 YSSSIDALINALAVLQLVLRALLSRFG-IAPALLIPGVALIMGYALA-----ASP 344
OY 385 LVAMFGTTPMLAVVGAIQNLSKSTKVALFDSTKEAAYPLDQEQVKGAALIDVVA 444
DB 345 L-----PMMLAITV---QVITRSSFPALAKAPARETLVTRVREMYKGAALIDTVVY 392
OY 445 R-----FKSGGALLIOGLVITCG-SIGA 467
DB 393 RGDPLTFWVWKLVSATGSS--AVFGVGLVATGTVGA 429

RESULT 21
O84237 PRELIMINARY; PRT; 928 AA.
ID O84237;
AC O84237;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C. PSITTACI hypothetical protein.
GN CT234.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Ck;
RC MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis";  
RT Science 282:754-759(1998).

```

DR EMBL: AE001296; AAC67826.1; --
 KW Complete proteome.
 SQ SEQUENCE 928 AA; 106121 MW; 8760C37E3D7CD75C CRC64;

Query Match 6.6%; Score 173.5; DB 16; Length 928;
 Best Local Similarity 19.9%; Pred. No. 0.00022;
 Matches 106; Conservative 88; Mismatches 195; Indels 143; Gaps 22;

7 KPGGLRS--FLMPHTHEKLVLMPELMFCITRYTVLRDTKTLVYGAGS----- 58
 6 KGGSLRLRLRFLLSLRGEEKRL-LFLLGLI-----WSVACYSGLALGSS 51
 59 -----GAELPPE-----IKFVLVPCALIFMLIVAKLSNLSKQALFYAVGTPPLI----- 104
 52 LPLEISGAEKLPALVIGASFL---CFISCLILVNLRSKRVSPKALFLSFICVLICNLX 108
 105 -FFALFPVLYPLRDVLPHTPEPADRLQALPPGLLGLVALILNNMTFAAFVYLAELMG-SV 162
 109 LFWHL-----AIH-----KVGSGTPFPL-----YRLI--IMGLTI 136
 163 MLSLMFGFANETIKHEAKRYALFGIGANISLSAGAIYMAK-LRASVSEGVDPNG 221
 137 LCVANFPGFIDQFNIQDAKRHCIF-----NATIFCGDFGARIVNOIQYLG 184
 222 ISRLIMMTIVSGVLMAASYWINKVLTDR-----FYNPEEMOKGKKAKPRMMK 275
 165 AELIILAFIVITTFPLVHYTSSLSKSEBHNLELDGYRPTKQ-----TLKCLK 238
 276 DSFLYLDSPYLLTLTLVIAVGICINLIEVTWKSQKLQYNNM--DYSEFMGNPSFWT 333
 229 DKYTYLVASFYLMQLLVFT-----EFNYLKI FDAQFGNAETCELTEFTYSSSWI 290
 334 GVVAVTLMLFVGAGNIRKFGMLTGALVTPVWVLLIGVFPALVIRNQGASGVANEGTTP 393
 291 SLGNPFALFAYSRVITRG-----INNITLPAICF-----FSLFCWS 330
 394 LMLAVVVAIQNILEKSTKVALFDSTKEMAYIPLDQEQVKGAADIVVAARFGSGAL 453
 331 IKTSTFIATMGIAEGALVADDDNNQLIYGINRKIRNQRIAE-----SF 379
 454 IQQGLLVTCISIGAMTPYLAIVLLFTIAMVSA---TKNKLFLAQSALK 501
 380 VEPAGMFICALCLFIPHOYVLCIISAVCILLAILLRTHYSKAILRNLISLE 431

RESULT 22

Q8NU46 PRELIMINARY; PRT; 538 AA.

AC Q8NU46; PRELIMINARY; PRT; 538 AA.
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative P-aminobenzoate transporter.
 GN CG10101.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 CC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 CX NCBI_Taxid=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RT Submitted (May-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AP005274; BAB97494.1; --
 DR InterPro; IPR004697; Put transpt_Abgt.
 DR Pfam; PF03806; Abg_transport; 1.
 KW Complete proteome: Hypothetical protein.
 SQ SEQUENCE 538 AA; 57353 MW; 9D0E4165E7646B4 CRC64;

Query Match 6.2%; Score 162; DB 16; Length 538;
 Best Local Similarity 21.1%; Pred. No. 0.00083;
 Matches 127; Conservative 83; Mismatches 210; Indels 182; Gaps 29;

QY 2 TKTEEK-----PFGKLR-----SPLMPHT--HELKVLMPMLMFCITRYTVLRDTKDT 50
 6 SKTSEKQPDAPPTSKLSKMSDKFLANGVETTLGNKLPPTFTLLILFLITALLASSSIAMNNV 65
 51 LIVAGGS-----GAELPPIKFLVPCAL-----IFM 79
 66 SVI-VPGSDEELFVKGFLPTGGLTWLTNGLANVIGFPPPLTLVPIILAVGVAERSGMLA 124
 80 LIYAKLSNLSKQALFYAVG---TPPLIFALPPTVLYPLRDV-----HPTFPARL 129
 125 ALIRLFGSAKKIVLPVAVGVIGVASTIMADAF-VVVPPLAAVFKAGRRPV----- 177
 130 QALPPGLLGLVALIRNMTFAAFVYLAELMGVSLMFGFANEI-----TKTHE 180
 178 -----AGLGSFAAV-----GAGYSTALV--PTSIDALFAGITNVAVMTLPGIATTEVNP 225
 181 AKRFYALFGIGANISL-LASGRAI-----VMASKLRASVSEGVDPNG-----ISRL 226
 226 VSNY--FNIASIVLGLCGFLIDKVLPEPMWOKIATEYAESIEPTSAADDEISATL 283
 227 -----LMAMTIVSGVLMAASYWINKVLTDRFNPPEEMOKGKKAKPR 271
 284 TAOENRALTISMWTLTALVAILLVVLLRSGSPMRNE-----GGLIPT 326
 272 MNMDSFLYLDSPYLLTLTLVIAVGICINLIEVTWKSQKLQYNNNDYSEFMGNFSF 331
 327 SPLSSVFT-----VLEPFMVGALYGNVGTIK-----NMDDVNNMGEAI- 369
 332 WTVGVSVLMLFVGAGNIRKFGMLTGALVTPVWVLLIGVFPALVIFENQAGL--VAMF 389
 370 -KDMIGFLVLAIFLIGQFVALFNM-----TGIGTWAV---QGAAGLEAIGLT 412
 390 GTPPLMAVVGALON--ILKSTKVALFDSTKEMAYIPLDQEQVKGAADIVVAARFG 447
 413 GPALIAFIILASCLNLIISGSAMWTL-----MAAFVMPFALGYEPFIQAAFRVG 466
 448 KSGGALI---OQGLLVTCISIGAMTP-----YLAIVLLFTIAMVSA---TKNKLFLAQS 498
 467 DSAQVITPLNPMYVILIGLRRYEPDAGLGLMSRLPIFPPLAMATILAIWTFYADL 526

QY 499 AL 500
 DB 527 PL 528

RESULT 23

Q9PKG0 PRELIMINARY; PRT; 918 AA.

AC Q9PKG0; PRELIMINARY; PRT; 918 AA.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical protein TC0505.
 GN TC0505.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CX NCBI_Taxid=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / N199;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Baes S.,
 RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL; AE002319; AAF39347.1; --
 DR TIGR; TC0505; --
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 918 AA; 105043 MW; 46D50A8FF198D887 CRC64;
 Query Match 6.1%; Score 161; DB 16; Length 918;
 Best Local Similarity 19.0%; Pred. No. 0.0018;
 Matches 88; Conservative 80; Mismatches 198; Indels 96; Gaps 14;
 QY 57 GSGAEAIPIFKFWLVCALIFMLIYAK---LSNLIKQALFYVAGTFFLIFPALFPTVI 113
 DB 13 GEEKRALFLFLILGLVSVACGSLALGESLFEISGTRKLPFAYLGSFLFC-ISCVL 71
 QY 114 YPLRDVLPTEPADRLQALIPGLGLVALIRNMTFAFYVLA-----EL 158
 DB 72 YSL-----SRKASPRALFLSPISCVLCNVLFWLAIHRGVSPTFLYRILI 121
 QY 159 WG-SVWLSLFWGANEITKHKRYALFGIGANISLASGAIYVASK-LRASVSEG 216
 DB 122 WGLTILCYANFWGVDPDFNIODAKRHFCIF-----NAITFCGDFLGARLVNQ 169
 QY 217 VDPWGISLRLMAMTIVSGVLVMAVYMNKNVLTDDR-----FYNPEMOKGKGAAP 270
 DB 170 IQHIGAEILILSFAVIVFTFPLVHYTSSLSKEISEHDLELDYGYPSAKQAP----- 223
 QY 271 KMNKMSDFLYLDRSPYILLLTLVYAGICINLIEVTWSQKLQYPMNDY--SEEMGN 328
 DB 224 KLCIKDKYTFYVSYFYLQQLVFT-----EFNYLKI FDAQFGNAKPFELTENPTK 275
 QY 329 FSEFTGVSVILIMLFVGGNVYRKFGMLTGALVTPVMVLIGIVFPALVFRNQSGLVAM 388
 DB 276 YSSWISLGNMFALFAVSRYVITRG-----INNIIIFADICF-----FSL 315
 QY 389 FGTTPLMLAVVGAIQVILSKSTKYALFDSKEMAYIPLODEQKVKKAIDVVAARFGK 448
 DB 316 FVCKMSITSVIATWGMARGLAVLDNNLQLLIYIGIPKNINQVRIALE----- 367
 QY 449 SGGALIOGGLVTCGSIGAMTPYLAIVLLFPIATMVSATRL 490
 DB 368 ---SFVEPAGMFCALCLFPHQYIICIIISAVCILLALL 406
 RESULT 24
 Q928P5 PRELIMINARY; PRT; 925 AA.
 AC Q928P5;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CT234 hypothetical protein.
 GN CPN0293.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
 RL Nat. Genet. 21:385-389(1999).
 DR EMBL: AE001614; AAD18442.1; -
 KW Complete proteome.
 SQ SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;
 Query Match 6.0%; Score 158.5; DB 16; Length 925;
 Best Local Similarity 20.6%; Pred. No. 0.0027;
 Matches 95; Conservative 94; Mismatches 163; Indels 109; Gaps 26;
 QY 59 GAEAIPIFKFWL---VPCAIIFMLIYAKLSNLSKQALF-YAVGTFFLIFPALFPTVIY 114
 DB 49 GSAELP--KTYLGSSLLICVLSLILYNLFKRIHSATRALFIPVSLICNCFYILSSIF 106
 QY 115 PLRDVLPTEPADRLQALIPGLGLVALIRNMTFAFYVLAELMGVMSL-LMFWGFRAN 173

DB 107 -----AIDPE-----RSLPF--FYRI-IMSLTILSYTSMGCFVD 138
 QY 174 EITKIHAKRFYALFGIGANISLASGRALVMAASKLRASVSEGVDPKGISLRLMAMTIV 233
 DB 139 QFNLDGKHNFICIFN-----AIIIFGDAL--OSGIIASLVHTIGIGIILLFTAAVLVT 191
 QY 234 SGLVMAVYMNKNV--LTDPR--FYNPEMOKGKGAAPKKNMKSDFLYLDRSPYIL 288
 DB 192 FPLV-----FYVSKLSKLSDDHDLFID-----TGHPPLSKALKCFYDKTYTYLL 238
 QY 289 -----LTLVLVYAGICINLIEVTWSQKLQYPMNDY--SEMGNSFTGVSVVILMT 342
 DB 239 CFYFLMQDLAIA-----TEFNVLKIFEIQFASKEEFELVAHIGKSLWISGNMCFAL 291
 QY 343 FVGGNVYRKFGMLTGALVTPVMVLIGIVFPALVFRNQSGLVAMGCTPPLMLAVVGA 402
 DB 292 FAVSRVYKRLG-----VNIIILFAPLCFSL-----LFLWTRKTLTSLIVLA 333
 QY 403 IQNLSKSTKYALFDSKEMAY--IPLODEQKVKKAIDVVAARFGKSGALIQGLLV 460
 DB 334 M--VREGVTVYALDNNLQLLIYGV---NKIRNQ--KIVVESFIEPIGMVMS---L 382
 QY 461 ICGSIGAMTPYLAIVLLFPIATMVSATRLKLFQAQALK 501
 DB 383 IC--FLSSQGVVFCILISLIALTLVC---LVRSYVAKALIK 418

RESULT 25
 Q9J5I6 PRELIMINARY; PRT; 925 AA.
 ID Q9J5I6
 AC Q9J5I6;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
 DE CT234 hypothetical protein.
 GN CPJ0293 OR CP0465.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=2015025; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey B.K., Peterson J., Uitterlind T., Bery K., Baes S.,
 RA Linher K., Weidman J., Khouiri H., Craven B., Bowman C., Dodson R.,
 RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai T., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunita S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AP002208; AAR38302.1; -
 DR EMBL: AP002546; BAA98503.1; -
 DR TIGR: CP0465; -
 SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;
 Query Match 6.0%; Score 157.5; DB 16; Length 925;
 Best Local Similarity 20.4%; Pred. No. 0.0032;
 Matches 94; Conservative 95; Mismatches 163; Indels 109; Gaps 26;
 QY 59 GAEAIPIFKFWL---VPCAIIFMLIYAKLSNLSKQALF-YAVGTFFLIFPALFPTVIY 114
 DB 49 GSAELP--KTYLGSSLLICVLSLILYNLFKRIHSATRALFIPVSLICNCFYILSSIF 106

```

Oy 115 PLRDVLTHTPEBADRLOALIPPELLGLVALIRNMTFAFYVLAELMGVSMLS-LMEWGFAN 173
Db 107 -----AIDPF-----RSPLEF-FYRIV-1MSLITLSTFSGFVD 138
Oy 174 EITKIHEAKRYVALFGIGANISLASGRAIYVASKLRASVEGVDPMGISRLMAMTIV 233
Db 139 QPFNLQDKRHFICIFN-----AIIPLGDAI--GSGIISLVHTIGIOGILLIFTLAVALT 191
Oy 234 SGLVIMASVWMTNKV--LTDP--FYNPEMOKKCAKPRMNMKDSFLVDSRPLYL- 288
Db 192 PPIV-----FYVSKSLKSLSDHDLFID-----THPPPLSKALKLCEYDKTFYLL 238
Oy 289 -----LTLVLVAYGICINLIEVTWKSQKLOYPNNNDYS--EFMGNFSPMTGVSVLIML 342
Db 239 CFYFLMQLALAI-----TEFNLIKIFEIOPASKEEFELVAHIKCSLMTSLGMCPL 291
Oy 343 FVGGNVIRKFGMLTGAULTVPMVLLTGIVFPALVIFRNQASGLVAMFSTPLMLAVVGA 402
Db 292 FAYSRIVKRLG-----VNNIILFAPLCFLS-----LFLFWTFKTLISIAVLA 333
Oy 403 IONLSKSTKXALPSTKEMAY--IPLDQOKVKGKAIADVAAAFGSGGALLIOGGLV 460
Db 334 M--VREBQVYALDNNLQLLIYGP--NKIRNQ--IRIVSESPLEPIGLVWS--L 382
Oy 461 ICGSIGAMPYLAVALIFLIIAIVLSATKLNKFLAQSALK 501
Db 383 VC--FLSSQOYVFCIIISLITILVC--LVRSYAKALK 418

```

RESULT 26

O8YNE4 AC O8YNE4: PRELIMINARY; PRT; 1002 AA.

```

DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein A114622.
GN A114622.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matsumoto A., Iizuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Nakazaki N., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazawa M., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Rep. 8:205-213(2001).
DR EMBL; AP003597; BAB76321.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1002 AA; 113795 MW; 4E060D76A2388BAE CRC64;

```

Query Match 5.7%; Score 151; DB 16; Length 1002;

Best Local Similarity 20.2%; Pred. No. 0.011; Mismatches 196; Indels 114; Gaps 23;

Matches 101; Conservative 88; Mismatches 196; Indels 114; Gaps 23;

```

Oy 19 IHTHELKVLPMFPCITFNVTYVLRDPTKTLIVGAPSGABAIPIFKFWLVVPCAIIF 78
Db 24 LRPEGERITWMMFAFYTVSGLRKABDSYVALPLDEGAG--PLP-----WMTYASAVMG 77
Oy 79 M--LIYAKLSNLSKQALFYAVGTPFLIFPALPEPTVYIPRDVLTHTPEADRLOALIPP 135
Db 78 MALFVVSWMLOKIPPLR-----WVVAIAIPCMIVPL-----ILLVLLRW 116
Oy 136 GL-----LGLVALIRNMTFAFYVLAELMGVSMLS-LMEWGFANITKIHEAKRYVALFG 189
Db 117 GIDVAVYSVIVVFLRLMV--DSIYVNDLNTSIV-----ANQLFNIRIKKTYPLIS 167
Oy 190 IGANISLASGRAIYVMA--SKLRASVSEGVDPMGISRLM--AMTIVSGVLVMSY 242

```

```

Db 168 SGLVADVIVSGFSLPMLLEFAKLN-----RYIMACGYIMIGSALICYLSY 213
Oy 243 WINKANVLTDPFRFNPPEMOKKGAKPXNMMDSPILDRSPYLLITLVLVAYGICIN 302
Db 214 QYRTSPFAPORLIPQOASRRRIQAP-----LKR--YTLQLFVGLQITIGL 261
Oy 303 LIEVTWKSQKLOYPNNND--YSEFMGNFSPMTGVSVLIMLFPYGVNIRKFGMLTGAUV 360
Db 262 LVDFQYIQLKTI--NIGDELAGFLGIFGIVGLCELTQWPFSSSLIERFGVFTFAL 318
Oy 361 TPVWVLLTGIVFPALVIFRNQASGLVAMFSTPLMLAVVGAII-ONLSKSTKXALPDST 419
Db 319 LPVAV--GKVVPQMTIVLILPQIGS-----LAFPMGLVGLKFEDELRTYFVSS 367
Oy 420 KENAYIPLDQOKKGAALDVVAARFGSGGALLIOGGLVIGSIGAMPYLAVALIFL 479
Db 368 GPLLYQPI--PERLRSTQV-----LSGGT--AEAIATGTAG--IIIVITLV 409
Oy 480 IAIWLSATKLNKFLAQS 498
Db 410 GGLF-VPAIMQKVFISGT 427

```

RESULT 27

O8TMZ1 AC O8TMZ1: PRELIMINARY; PRT; 533 AA.

```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Transmembrane efflux protein.
GN VARS OR MA2503.
OS Methanobacteria aceticivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanobacteriales;
OC Methanobacteriaceae; Methanobacterina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atchao D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwen P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grisham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanobacterina aceticivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010943; AAM05889.1;
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR Pfam; PF00083; Sugar_tr_1.
DR PROSITE; PSS0850; MFS; 1.
KW Complete proteome.
SQ SEQUENCE 533 AA; 56009 MW; A10C15C77FB89A34 CRC64;

```

Query Match 5.6%; Score 147; DB 17; Length 533;

Best Local Similarity 19.3%; Pred. No. 0.01; Mismatches 184; Indels 192; Gaps 21;

Matches 114; Conservative 83; Mismatches 184; Indels 192; Gaps 21;

```

Oy 23 ELKVLV-PMFMEFCITFNVTYVLR-----DKRDTLIVGAPSGABAIPIFKFWLV 71
Db 2 ERKTVLVALVLANFVLVIDTTIMVNSISALISPNNTVTVQQA-----ITLYAL 51
Oy 72 VPCAIIFMLIYAKLSNLSKQALF-----YAVGTPFLIFPALPEPTV----- 112

```

Db 52 VWAS--FMITGGIGDIIIGKRRAPRLGVIYGVGS---LLTAISPTIGVLFWGMSILEGL 106
 Qy 113 ----IYPLRDVHPTEPDRLOAIIIPGLGLVALIKMTFAFYLAELMSWLSLMP 168
 Db 107 GATLWPAIQTIVTNYEGEDRAI-AYGIIGIV-----ASGIALGPIIGGMLTTAVT 158
 Qy 169 W--GFANEITIKHEAKRFYALFGIGANISILASGRALVWASKLRASVSEGVDPWGISLRL 226
 Db 159 WRLNAPGEVIVY-----IIVLALSRLIDA-PLKAGEBKLVIGSILSA 202
 Qy 227 LMAMTVSGVLVMAWY-WINKNVLTDPFRYNEPMQKKGAKPRMKNDSFLYLDSP 285
 Db 203 LGMGLIVFGIIMAGTYGWW-----KARQTFSIAGIEISIP 236
 Qy 286 YLLLLTLVIAAGICINLEVTWKSQKLQYPMNDYSEFMGNFSGFWGVSVLIMLPFG 345
 Db 237 FGLSPFPVFIAGVILIGFAAMERHL-----ISSGMPPLVR 273
 Qy 346 GNVRKFGMLTGLATVPVWVLLTGIVFPALVIFRNOASGLVAM---FGTPIMLAVVGA 402
 Db 274 LDVLRDRVTSGLIQMWQTLFGGFLPSMALFLQIVGLANMQTFYTLPLSTPLIAS 333
 Qy 403 IONILSKSTKXALPDSTK-----MAYIPLDOE----- 430
 Db 334 L-----TASRLSIFASKRIIQAGLVILLAGFLAIFATIDVEVRGLGMLTGALIGIGG 388
 Qy 431 -----QYKGAALDVVAARSGSGGALIOGGLVIGSISAMPIYLA 474
 Db 389 LIASQVMNLVLSQVTPERISSETALMGTSQNLGAMIGTALMGLLAVAGAVAIT----- 443
 Qy 475 ILFLFIAT-----WLSATKLNKFLAQSALK 501
 Db 444 LIDSTALPEDLKPDLVSAVENQFLSNEIQ 476

RESULT 28

Q9KSH7 PRELIMINARY; PRT; 540 AA.

AC 09KSH7, 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Transporter, BCCF family.
 GN VC1329
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El TOR N16961 / Serotype O1;
 RK MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolenko M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.,
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae."
 RL Nature 406:477-483(2000).
 DR EMBL; AE004207; AAF94438.1; --
 DR TIGR; VC1279; --
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF; 1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TIGRFAMs; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Complete proteome.
 SQ SEQUENCE 540 AA; 58464 MW; DAA6C5A49EE6B0E CRC64;

Query Match

5.5%; Score 145; DB 16; Length 540;

Best Local Similarity 20.5%; Pred. No. 0.015;
 Matches 106; Conservative 76; Mismatches 187; Indels 148; Gaps 24;

Qy 69 WLVPICAIIPMLI-----YAKL-----SNILSQALFYAVGTPP-LIFPALF 109
 Db 79 WLFWSGNLFLFLFCVGLVISPFGKIRLGQAQKADHSFLSWLAMPFAAGMGIGLFWWSVA 138
 Qy 110 PTVY-----PLRDVHPTEPDRLOAIIIPGL-LGLVALIKMTFAFYLAELMS 161
 Db 139 EPVAYTGWETPL-----GVEANTPQAQLMGATMFGHLPW---AIXGV 183
 Qy 162 WLSLMP-----MGFANEITIKHEAKRFYALFGIGANISL-- 196
 Db 184 VALSLAFPCYNKGLPELSRSVFYPLGGRANGACHVVDILAV--LATLFGALTSGLGA 241
 Qy 197 -LASGRALVWASKLRASVSEGVDPWGISLRLMAMTVSGVLVMAWYINKNVLTDP 254
 Db 242 QQAASGHHVFGFES-----GIGQIAV-ICVVTGLATLSVRGID----- 281
 Qy 255 FYNEPMQKKGAKPRMKNDSFLYLDSPFIYLLTLVIAAGICINLEVTWKSQKL 314
 Db 282 -----GGVXVISNIMTVAFL-----LLALAVALGYSTVLTLSIPTTMAVILON 324
 Qy 315 QYPMNNDYSR-----FMGNFSF-----WTGVSVLIMLPVGNVRKFGMLTGLATVP 362
 Db 325 LIPLSNPHGRDEWVFGQWTFYFAMWISNSPFGMFAVRSKRTIRF--IYAVLILP 382
 Qy 363 VVWLLTGI-VFPALVIFR--NOASGLVAMFGTTPMLA-----VVVGAIONILSKSTK 412
 Db 383 TLVTLVWMSVFGGLAVDPQVINEI-GVLQNGGLTDSVLSLPMQFDBELVGKVLVAIVLV 441
 Qy 413 YALPDSTEMAVIPDO-----EQYKGAALDVVAARSGSGGALIOGGLVIT 461
 Db 442 LVFFITSSDGLVIDSITSGKLDSPVLORFYNAVILGGAVALVALLMIGTEAVQALO-- 499

RESULT 29

Q9JUS7 PRELIMINARY; PRT; 660 AA.

AC 09JUS7, 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Putative transmembrane transport protein.
 GN NMA1483
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RK MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.,
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491."
 RL Nature 404:502-506(2000).
 DR EMBL; AL162756; CAB84716.1; --
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF; 1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TIGRFAMs; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Transmembrane; Complete proteome.

SQ SEQUENCE 660 AA; 74418 MW; 50046CF6DA3A37D1 CRC64;
 Query Match 5.5%; Score 144; DB 16; Length 660;
 Best Local Similarity 20.7%; Pred. No. 0.022;
 Matches 127; Conservative 99; Mismatches 171; Indels 216; Gaps 34;

QY 32 LMFPCITFNVTYLRDTKDTLIVGAPG-----SGAEAFPIKF-WLVPCALIFM----- 79
 DB 7 LFFVCVIV-----VLVLTVPDQVOMLDRAKEVIFTEFSFVLTSTIFLGLLI 56
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFPTVIYPLRDVLPTEF 125
 DB 57 LSVSSLGIRIGRDEDEVEFGFLSWLMLPFAAGMGVGLMFVGAEPMLHMFSDITAGTPE 116
 QY 126 ADRLQALIPGLIGLVALILRNWTFPAFYVLAELMGVSLMPFGFANEITIKHEAKR-- 183
 DB 117 HROQOALL-----HTVFHWGVHAMSV---YGTIALALAYFGFR---YKLPALRSC 161
 QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVSE--GVDPWGISLRLL 227
 DB 162 FIPPLKEXISGRFGAIDIMALLATFPGIITTLGFGASQDGLQDGIEMIAENSFSVOYL 221
 QY 228 MAMTIVSGVLVMAASYWMINKVLTDPFRFYNPEMOKKKGAK--PKMMKDSFL---YL 281
 DB 222 IIAAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLLFFVL 261
 QY 282 DRSPYITLLTLVIAVYICL-----NLIEVTWKSQKLQYPMNDYSEFMGNFSF----- 331
 DB 262 AAGPVTYLLS-----AFGDNIGNYLGVLRLSFKT-----YAYEREHKPFESMTVLVYAW 312
 QY 332 ---WTGVSVLIMLFVGNVIRKFGMLTGALTVPVWVLTGLVFPALVIFRQASGLVAM 388
 DB 313 WCSNAPFVGLFPIARISKRTIREF--VFGVLLIPG--LFGVLWF-----TV 354
 QY 389 FGTTPLML-AVYVGAIONILSKSTKYLFDSTKEMAYIPLDOEKV----- 433
 DB 355 FGNTAIWINDGVAAGVLEKMTSSPETLIF--KFNFYLPPLPELTSIVSLVLSLFFVYSA 411
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIQGLLVICSGIGAM 468
 DB 412 DSGIYVLNNITSRDGLSAPRMQAVMMGVLMGSAVAVLLMRSGG-----LGNLQSM 461
 QY 469 T-----PYLAVILFLIAIWL-VSA-----TKLN-----KLFLAQSALKEQ 503
 DB 462 TLIVSLPPLMLLMICFSLMKGLSADKKYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 521
 QY 504 EV---AQEDSAPA 513
 DB 522 DILKFLKHTASPA 534

RESULT 30
 Q9JZ61 PRELIMINARY; PRT; 675 AA.
 AC Q9JZ61;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-UN-2002 (TREMblrel. 21, Last annotation update)
 DE Transporter, BCCF family.
 GN NMB1277.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 CC Neisseriaceae; Neisseria.
 OK NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tetteijn H., Saunders N.-J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Hickey E.K.,
 Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Dickson R.J.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,

RA Cotton M.D., Utlarback T.R., Khouri H., Qin H., Vamathavan J.,
 Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
 Smith H.O., Frazer C.M., Koxon E.R., Rappuoli R., Venter J.C.,
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 RT MC58."
 RL Science 287:1809-1815 (2000).
 DR EMBL; AE002476; AAF1653.1; -.
 DR TIGR; NMB1277; -.
 DR InterPro; IPR000060; BCCF_transporter.
 DR Pfam; PF02028; BCCF; 1.
 DR ProDom; PD01011; BCCF_transporter; 1.
 DR TRIGRAMS; TIGR00842; bccf; 1.
 DR PROSITE; PS01303; BCCF; 1.
 KW Complete proteome.
 SQ SEQUENCE 675 AA; 76283 MW; 4BEC798C51D40196 CRC64;

Query Match 5.5%; Score 144; DB 16; Length 675;
 Best Local Similarity 20.7%; Pred. No. 0.022;
 Matches 127; Conservative 100; Mismatches 170; Indels 216; Gaps 34;

QY 32 LMFPCITFNVTYLRDTKDTLIVGAPG-----SGAEAFPIKF-WLVPCALIFM----- 79
 DB 22 LFFVCVIV-----VLVLTVPDQVOMLDRAKEVIFTEFSFVLTSTIFLGLLI 71
 QY 80 LIYAKLSNI-----LSKQALFYAVGTFP-LIFPALFPTVIYPLRDVLPTEF 125
 DB 72 LSVSSLGIRIGRDEDEVEFGFLSWLMLPFAAGMGVGLMFVGAEPMLHMFSDITAGTPE 131
 QY 126 ADRLQALIPGLIGLVALILRNWTFPAFYVLAELMGVSLMPFGFANEITIKHEAKR-- 183
 DB 132 HROQOALL-----HTVFHWGVHAMSV---YGTIALALAYFGFR---YKLPALRSC 176
 QY 184 FYAL-----FGIGANI-SLASGRAIV-----WASKLRASVSE--GVDPWGISLRLL 227
 DB 177 FYPPLKEXISGRFGAIDIMALLATFPGIITTLGFGASQDGLQDGIEMIAENSFSVOYL 236
 QY 228 MAMTIVSGVLVMAASYWMINKVLTDPFRFYNPEMOKKKGAK--PKMMKDSFL---YL 281
 DB 237 IIAAVMSLAIVSA-----ISGVKGKVKVLSLNLGLAFLLFFVL 276
 QY 282 DRSPYITLLTLVIAVYICL-----NLIEVTWKSQKLQYPMNDYSEFMGNFSF----- 331
 DB 277 AAGPVTYLLS-----AFGDNIGNYLGVLRLSFKT-----YAYEREHKPFESMTVLVYAW 327
 QY 332 ---WTGVSVLIMLFVGNVIRKFGMLTGALTVPVWVLTGLVFPALVIFRQASGLVAM 388
 DB 328 WCSNAPFVGLFPIARISKRTIREF--VFGVLLIPG--LFGVLWF-----TV 369
 QY 389 FGTTPLML-AVYVGAIONILSKSTKYLFDSTKEMAYIPLDOEKV----- 433
 DB 370 FGNTAIWINDGVAAGVLEKMTSSPETLIF--KFNFYLPPLPELTSIVSLVLSLFFVYSA 426
 QY 434 -----KGKAA-----IDVVAARFGKSGGALIQGLLVICSGIGAM 468
 DB 427 DSGIYVLNNITSRDGLSAPRMQAVMMGVLMGSAVAVLLMRSGG-----LGNLQSM 476
 QY 469 T-----PYLAVILFLIAIWL-VSA-----TKLN-----KLFLAQSALKEQ 503
 DB 477 TLIVSLPPLMLLMICFSLMKGLSADKKYFETRVNPTSVFTGKWKERLVQIMSQTOEQ 536
 QY 504 EV---AQEDSAPA 513
 DB 537 DILKFLKHTASPA 549

RESULT 31
 Q8FUB5 PRELIMINARY; PRT; 563 AA.
 AC Q8FUB5;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Conserved hypothetical protein.

OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NC NCBL_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125;
 RX MEDLINE=99184645; PubMed=10086841;
 RA Takami H., Nakaone K., Hirama C., Takaki Y., Masui N., Fuji F.,
 RA Nakamura Y., Inoue A.,
 RT "An improved physical and genetic map of the genome of alkaliphilic
 RT Bacillus sp. C-125."
 RL Extremophiles 3:21-28(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL, AB013375; BAA75389.1; -
 DR InterPro; IPR004638; Efflux_EmrB.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub_transporter.
 DR InterPro; IPR014111; TCR_TcRb.
 DR Pfam; PF00083; sugar_trf.1.
 DR PRINTS; PRO1036; TCR_TcRb.
 DR TIGRFAMs; TIGR00711; efflux_EmrB.1.
 DR PROSITE; PS50850; MFS; 1.
 KM Transmembrane.
 SQ SEQUENCE 475 AA; 51336 MW; 1442844030A1B8E5 CRC64;

Query Match 5.3%; Score 140; DB 2; Length 475;
 Best Local Similarity 22.7%; Pred. No. 0.03; Indels 88; Gaps 22;
 Matches 110; Conservative 74; Mismatches 212;

QY 67 KFMVLVPCALIFMILYAKLSNLSKQAL-----FYAVGTPFLI-FFALPTVIYELR 117
 DB 7 KKMVVV-CAVIFGSPFTMLNNNSMNPALPOLMNVFEADAVATGVITTFVAMGMTPLE 65
 QY 118 DVLPHTPADRLQAILPPGLIGLVAIL--RWMTAAFYV---LAEIMGSV-----LSIME 168
 DB 66 GYL-GDKIKKEAYILGIGIFVLSLIGALSMLPFLVFRGLOGIGGGMPLSMTLIF 124
 QY 169 -----MGFANETK-----IHEAKRFYALPGIGANISILASGRALVW 205
 DB 125 DAFPRNERGLATGVWGVASMAPIPTGPTLPGFPIVETSNMKYLFVNIPFGLIGITAAVY 184
 QY 206 ASKLRASVEGDVPMGISL-----RLIMANTIVSGVLMAVWINKNVLTDPREFNPE 259
 DB 185 LPKIARSGQIKLDMNGFLFYAGVSGILLARMSDLTHLE--WINGVLLLLGLALCLIV 242
 QY 260 EMQKGKKGAKPKMMKDSFLYLDSPYLLILLIVAYGI-CINLIEVTWKSQKLYPN 318
 DB 243 FVQVEKRAEQPLLD-----LSLFRIP--AYSLSIWVAGISSIGIFAGIFVPLLDQ--Q 292
 QY 319 MNDSSEFMGNFSFMTGVSVLIMLFVGGNVIRKFGMLTGAALVTPVMV-----L 366
 DB 293 VYDGPMTGTGLVPLPSALFTGLTWSIGRMIDKRG-PSGINTAGMTAAVGTPALGYIHL 351
 QY 367 LTGIVF-FALVIFENQASGLVAMFGTPTPLMLAVVGAIONILSKSTKALPDSTKEMAYI 425
 DB 352 ETGIMYIFAMMAIIVGVGGLTTPPATTTGNMAIPEGILSRGSANNVLRQSSAFGIYFI 411
 QY 426 PLDDEQKVGKKAIDVVAARFGKSGGALLIQGLLVGSGISGAMTPYLAIVILLFIATLV 485
 DB 412 SVPEFVR-RQGLAL--VSTSFEEATLQAINEGFFI---VGFLT-----ALSDPAAYWLE 459
 QY 486 SATK 489
 DB 460 KKAK 463

RESULT 34
 020628 PRELIMINARY; PRT; 536 AA.
 AC 020628;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Cytochrome oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase
 DE polypeptide I).
 GN COXI.
 OS Cythecocodium cohnii (Dinoflagellate).
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Dinophyceae; Gonaulacales; Cythecocodiaceae;
 OC Cythecocodium.
 NC NCBL_TaxID=2866;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WH-d;
 RX MEDLINE=97424386; PubMed=9280308;
 RA Norman J.E., Gray M.W.;
 RT "The cytochrome oxidase subunit 1 gene (cox1) from the dinoflagellate,
 RT Cythecocodium cohnii".
 RL FEBS Lett. 413:333-338(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WH-d;
 RA Norman J.E., Gray M.W.;
 RT "A complex cox1 gene organization in the mitochondrial genome of the
 RT dinoflagellate, Cythecocodium cohnii: Homologous recombination
 RT generates two different open reading frames".
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME
 CC C + 2 H(2)O.
 CC -1- PATHWAY: RESPIRATORY CHAIN; TERMINAL STEP.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 DR EMBL; AF186994; AAB69658.1; -
 DR HSPSP; P18401; 1PFT.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; CYCOXIDASE1.
 DR PRINTS; PRO1165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Copper; Electron transport; Heme; Inner membrane; Membrane;
 KW Oxidoreductase; Respiratory chain; Transmembrane; Transport;
 KW Mitochondrion.
 SQ SEQUENCE 536 AA; 59328 MW; AA6BB4C821BD2C61 CRC64;

Query Match 5.2%; Score 137; DB 8; Length 536;
 Best Local Similarity 20.3%; Pred. No. 0.056;
 Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;

QY 68 FMIIVPCALIFMILYAKLSNLSKQALFYAVGTPFLI-FFALPTVIYPLDVLHPTPAD 127
 DB 81 FYLVMP-----GLFGGFGNYFIVIFGSGSEVVP----- 109
 QY 128 RLQ--ALPPGLGLVAIL-----RWMTF-----AAYVLA-ELMGSYMLSLFWG 170
 DB 110 RVNPFSLILSLSYLLILSLISEFGGGMWLYPPISTFPMILSPSSIGITLGLISG 169
 QY 171 FANEITKHEAKRFYALFGIGANISILASGRALVAMSKLRASVEGDVPMGI---SLRL 227
 DB 170 ISSVLTSLNFWITLISRSIGITLKITS-----LFPMSLITSGML 211
 QY 228 MAMTIVSGVLMAVWINKNVLTDPREFNPEMOKKKGAKPKMMKDSFLYLDSPY 286
 DB 212 LTLPILTGALLMI-LSDINVTLLFFDPIF-----GDPPIFYQHLFWFGHPEVY 259
 QY 287 ILL-----TLIVAYGI-----CINLI-EVTWKSQKLYPNMNDSEF 325
 DB 260 ILIIPAGIISITISGLQKIFGNPSPMIFAMCSISLGSVWGHMYTTLGLETDTSYF 319
 QY 326 MGNFSFMTGVSVLIMLFVGGNVIRKFGMLTGAALVTPVMVLT-----GIVFALVIFR 379


```

Db 320 SG-----VTLLISLPTGKTI---FNNLSTYLGNSLKLKTNSSLFGILLFL-LMFTI 367
Qy 380 NQASGLVAMGCTTPMLAV-----VGAIONILSKSTKVALPDSKEMAYIPLDQEQVK 434
Db 368 GGSSTGII--IGNAAVDGLHDYTYIIAHFHFVLSLGAIVAFSG-----IIFNIKTIIG 419
Qy 435 GKAAIDVVAARFGSGGALIQOGLVLCG-----STG 466
Db 420 SKNLPSCSSNNRNVLVFLITGILITFGPMHFLGFVMPRRIRIPDPDSFISNFFLSIG 479
Qy 467 AMTPYLAIVLLFIILVLSATKLNKFLAOSA 499
Db 480 SGITLLSFGFLFKNCWITSLAOCSSRLFLGHVA 512

RESULT 35
Qy 094Y38 PRELIMINARY; PRT; 577 AA.
ID 094Y38;
AC 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Cytochrome oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I).
OS Cythecodinium cohnii (Dinoflagellate).
OC Mitochondrion.
OC Burkaryota; Alveolata; Dinophyceae; Gonyaulacales; Cryptecodiniaceae;
OC Cryptecodinium.
OX NCBI_TaxId=2866;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WH-d;
RX MEDLINE=21531838; PubMed=11675595;
RA Norman J.E., Gray M.W.;
RT "A complex organization of the gene encoding cytochrome oxidase subunit 1 in the mitochondrial genome of the dinoflagellate, Cryptecodinium cohnii: Homologous recombination generates two different cox1 open reading frames."
RL J. Mol. Evol. 53:351-363(2001).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICTOCHROME C + 2 H(2)O.
CC -1- PATHWAY: RESPIRATORY CHAIN; TERMINAL STEP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
EMBL: AF182643; A011630.1; -
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.1.
KW Copper; Electron transport; Heme; Inner membrane; Membrane; Mitochondrion; Respiratory chain; Transmembrane; Transport;
SQ SEQUENCE 577 AA; 64144 MW; FB8955EADCE78BFC CRC64;

Query Match 5.2%; Score 137; DB 8; Length 577;
Best Local Similarity 20.3%; Pred. No. 0.061;
Matches 104; Conservative 73; Mismatches 174; Indels 162; Gaps 22;
Qy 68 FMUVPACIIIFMLIYAKSNLSKQALFYAVGTPFLIPFALFPYVIYPLRDVLAHPTPAD 127
Db 122 FYLVMP-----GLFGGNGFYIYFQSGPEVVP----- 150
Qy 128 RLQ--ALLPGLGLVAIL-----RNMVF-----AAFYVLA-ELMGSVMSLMPWG 170

```

```

Db 151 RVNPFSLILSLSLVLLILSLISEFGGCTGWTLYPPLSTFPMNLSPSSIGNIIFGLISG 210
Qy 171 FANEITKIEHAKFPYALFGIGANISLASGRAIYVASKLRASVSEGVDPGCI--SLRL 227
Db 211 ISSVLISLNFWITILSLRSIGITLKTIS-----LFPMSLITSGML 252
Qy 228 MANTIVSGVLMSYMMINKVLTDPREFNPEEMQKGGKAKPKMMKOSFLYLDSPY 286
Db 253 LTPILTLGALLMT-LSDINYNLTLPDPIF-----GGDIFQHLFWFGHPEVY 300
Qy 287 ILVL-----TLVIAVGI-----CINLI-EVTKSQDLQYPMNDYSEF 325
Db 301 ILIIPAFGIISIIISGLQKIIIFGNPSMTFAMCISILSGVGMHHYITGLFETDSYF 360
Qy 326 MGNFSEWTVGSVLMLFVGANVIRKFGULTGALVTPVMLT-----GIVFALYIFR 379
Db 361 SG-----VTLLISLPTGKTI---FNNLSTYLGNSLKLKTNSSLFGILLFL-LMFTI 408
Qy 380 NQASGLVAMGCTTPMLAV-----VGAIONILSKSTKVALPDSKEMAYIPLDQEQVK 434
Db 409 GGSSTGII--IGNAAVDGLHDYTYIIAHFHFVLSLGAIVAFSG-----IIFNIKTIIG 460
Qy 435 GKAAIDVVAARFGSGGALIQOGLVLCG-----STG 466
Db 461 SKNLPSCSSNNRNVLVFLITGILITFGPMHFLGFVMPRRIRIPDPDSFISNFFLSIG 520
Qy 467 AMTPYLAIVLLFIILVLSATKLNKFLAOSA 499
Db 521 SGITLLSFGFLFKNCWITSLAOCSSRLFLGHVA 553

RESULT 36
Qy 033750 PRELIMINARY; PRT; 498 AA.
ID 033750;
AC 033750;
DT 01-JUN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN NDH3.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxId=32049;
RN [1]
RP SEQUENCE FROM N.A.
RA Klughammer B., Sulemeyer D., Badger W.R., Price G.D.;
RT "Involvement of ndh3, ndh3 and ORP427 genes in high affinity CO2 uptake in the marine cyanobacterium, Synechococcus sp. strain PCC7002."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
EMBL: U97516; AAB62185.1; -
DR InterPro: IPR003918; NADHdb Oxred4.
DR InterPro: IPR001750; Oxidored q1.
DR Pfam: PF00361; oxidored q1.1.
DR PRINTS: PR01437; NDOXDRDTHSE4.
KW NAD; Oxidoreductase; Plastocyanine.
SQ SEQUENCE 498 AA; 53900 MW; A0FEB032A4A8DE CRC64;

Query Match 5.2%; Score 136.5; DB 2; Length 498;
Best Local Similarity 20.6%; Pred. No. 0.056;
Matches 92; Conservative 60; Mismatches 139; Indels 155; Gaps 20;
Qy 82 YAKSNLSKQALFYAVGTPFLIPFALFPYVIYPLRDVLAHPTPADRLQALP 135
Db 55 YTFEHWMLSTIGLNYNVLGVDGLSLPLIVNSLTLVLAIVYIGSNHPRKLYSLILINS 114
Qy 136 GLIGVAIILNNTPAFAFYVLAELMGSVMSLMPWGFANETTKHAEKFAALRGIGNIS 195
Db 115 GIGG--ALIAN-----NLLFLFLY--EL-----ELIPFLLLAI----- 145
Qy 196 LLASGRAIVASKLRASVSEGVDPWGISRLMLAMNTVSGVLMASYW--WINKVLTLD 252

```

```

Db 146 -----W-----GGEKKGYASTKFLIYTAISGLVLAFLGIWVLSQSSND 186
QY 253 PRFINPEEMOKGKAKPRKMMKDSFLYLDSPYLLTLTLVYAGICINLIEV-TKMSQ 311
Db 187 FENLTLENIERTK-----VILLTILLIGIGIKIPVPLHTW--- 223
QY 312 LKLOYPNNDYSEFMGNFSFWTGVVSVLIMEFVGNNVIRKFGMLGALVTVPVLLTGIV 371
Db 224 LPDAVEANP-----AVTLL-----GGVPAKIG-----TYGLV 252
QY 372 FFALVIFR--NQASGLVAFMGTTPMLAVVGAIONILSKTXKALFDSKEMAYIPLD 428
Db 253 RFGQLPDPVMSVSPALAVIGTVSMVGLAALIQRLDKRVAYS---SIGHMGYILVS 309
QY 429 Q-----EKKYKGAIDVVAARFKSGGALLIQ 456
Db 310 TAACTELSLGAVACMISHLALLFHLVGIIEKRV-GRDLVDLNGMNPVRGLPLTS 368
QY 457 GLVVGSGIGAMTP---YLAVILLE 478
Db 369 SLTIAGMASGIPGLVGVFAEFLVF 394

```

RESULT 37

```

ID 08FH20 PRELIMINARY; PRT; 465 AA.
AC 08FH20.
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Nitrite extrusion protein 1.
GN NARK OR C1684.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=21792;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / ATCC 700928;
RX MEDLINE=22386234; Pubmed=1471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Domeneberg M.S., Batters F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016760; AAN80151.1; -.
KW Complete proteome.
SQ SEQUENCE 465 AA; 49877 MW; 73683DDB64CEFF5 CRC64;

```

Query Match 5.2%; Score 135.5; DB 16; Length 465;
 Best Local Similarity 22.5%; Pred. No. 0.062;
 Matches 109; Conservative 68; Mismatches 174; Indels 133; Gaps 28;

```

QY 69 WLVPVCAIIFMLIAKSNL-LSKQALFYAVGTPFLIFALFPYTYPLRVLHPTERAD 127
Db 41 WISPCLLIACVWMLSAVANLPEKGFNTTDLLEMLTALPSVSGALLAV--PYSF-- 96
QY 128 RLQALPLPGLGLVAIL--RNMTPAFVYLAELMGVSLMFMGFANEITKIHAKRFY 185
Db 97 -----MVPFGGRMT--AFST-----GILLIPCVWJGFA-----VDSTIRPY 132
QY 186 ALFGIGANISLASGRAIWMASKLRASVS-----EGVDPWGISLRLLMAM 230
Db 133 SVFII--ISLLCGFAGANFASSM-ANISFFPKQKOGALGLNGGNGMGSVMQVLAP 188
QY 231 TIYGVGLVMASVYMWINKVLLDPRFYNPEEMOKGKAKPRKMMKDSFLYLDSPYLL-- 288
Db 189 LVVS-LSPFAVF-----SQGVKQPDGTE--LYLANASWITVP 223
QY 289 LTLTLVYAGICINLIEVTWKSQLQYPMND-----YSEFMGNF-SFWTGV--- 335

```

```

Db 224 LLAIFTIAWEGMNDL-ATSKASIKEDLPYLKRGHLMWSLLYLATFGSFIGSAGEFAML 282
QY 336 -----VSYLIMFVG---GNVIRKFGW-----LTGALTVPV---MVLITGVFPALV 376
Db 283 SKTQFPDVOILQVAFSPFIFGALARSAGALSDRLGSTRVYLVNFIIMAFSGLLFTLP 342
QY 377 IFRQASGLVAFMGTTPMLAVV-----GALONISKSTKYALFDSKEMAYIPLD 430
Db 343 T-DQGGSPFAAFPA--VFLLPLITLAGSGSTFGOMSVIFRKLTMRYAEG---GSD 394
QY 431 QKVKGAIDVVA-----RFGSGALLIOGL--LVIGGS-IGAMTPYLAIVILLE 481
Db 395 ERARERATDTAALGFIISAIGAGFFIPKAFSGSIALTSGSPVGMKVFILFIACVVI 454
QY 482 INLV 485
Db 455 TWAV 458

```

RESULT 38

```

ID 0916P7 PRELIMINARY; PRT; 441 AA.
AC 0916P7.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Probable MFS transporter.
GN PA0241.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou Y.,
RA Garber R.L., Goltzy L., Tolentino F., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A.G., Larbig K., Lim R.M.,
RA Smith K.A., Sailer M.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Senter D.H., Hancock R.E.W., Lory S., Olsson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004462; AAG03630.1; -.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub. transporter.
DR Pfam; PF00083; sugar tr. 1.
DR PROSITE; PS50850; MFS; 1.
KW Complete proteome.
SQ SEQUENCE 441 AA; 48704 MW; 72EDFF82986FB601 CRC64;

```

Query Match 5.1%; Score 135; DB 16; Length 441;
 Best Local Similarity 21.1%; Pred. No. 0.063;
 Matches 110; Conservative 58; Mismatches 163; Indels 190; Gaps 26;

```

QY 32 LMFPCITFNTVLRDTCYIVG-----APSGAEKIPF--IKFVLVPCALIFMLIYA 83
Db 39 LVFPAATLNYI-----DRAALGVMPILAEKMSWTMDYANINFWQVYAGFIQGR 92
QY 84 KLSNLSKQALFYAVGTPFLIFALFPYTYPLRVLHPTERADLQALIPGLGLVAI 143
Db 93 FIDKVGKRAFFLAIV-----LMSLATGHHGLATS 122
QY 144 LRMWTFAPFV-LAEI--MGSVMLSMFMGFANEITKIHAKRFY--LFGIGANISLA 198
Db 123 AAGFMVORFLIGLLEANVYACVKTIRLWEPAGE-----RAVANGINAGNNGAV 174
QY 199 SGRAI-----VMASKLRASVSEGVDPWGISLRLLMANTVSGVLASVYMWINKVITDP 253
Db 175 TPALLPLILGVWG-----WQAAFICMSAL--GLVWLV-FWVRN----- 209

```

```

QY 254 RFYNPEEMOKKKG-----AKPKNMKDSFLYDRSPYLLTLVLVIA-----296
D 210 -YNPBEPHRYKQSELEYIQOQEBPPATRVPSQILRRGMARFALASITAPVFWFLY 268
QY 297 -----YQICNLLEVTWKSQKLOYPNNNDYSEFMGNFMTGVSVLIML-----342
D 269 WLPPFLNOQYQIGISVTQM-----GIPLLILMLTADFG 301
QY 343 FVGNVIRKFGMLGALVTPVWVLLTGIVFAL-----VIFRNQASGL-VAMFGTTPMLA 397
D 302 SVGGGILS--SMLRGPMARLARLSMLLPACTIVGVAFANASGLMIIV-----LAIA 354
QY 398 VVVGAIQ-----NLSKSTKVALFDSTKEMAYIPLDQEQVKAKAIDVVAARFG-----KS 449
D 355 LAVGHHQWMTANMS-----LYMDYTPKHLMSVTVG-----FGCMCAL 393
QY 450 GGALIQGGLVTCGSIGAMTPYLAVILFIITMLVLSATKL 490
D 394 GGMEMTQ---IVGVLTATNNNAVAVLFTMIPAMYPFIALTWL 431

RESULT 39
ID Q8G4F6 PRELIMINARY; PRT; 404 AA.
AC Q8G4F6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Bli430.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22949477; PubMed=12381787;
RA Schell M.A., Karamantzou M., Snel B., Vilanova D., Berger B.,
RA Persi G., Zakhlen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
KM EMBL: AF014773; AAN5227.1;
KN Hypothetical protein; Complete proteome.
SQ SEQUENCE 404 AA; 43122 MW; A6D16CC67EC251B7 CRC64;

Query Match 5.1%; Score 134.5; DB 16; Length 404;
Best Local Similarity 20.9%; Pred. No. 0.062;
Matches 92; Conservative 74; Mismatches 168; Indels 107; Gaps 24;

QY 95 FYAVGTPPLFFALPFTVYPLRDVLPHTPEADRLQAILP-----PG 136
D 9 FISGLDPSILSGAMPTW---SODLNVAVAMAGISAVISMFTIVSALLSDRMTLKFGAG 65
QY 137 LLEGVAL-LRWTRAFYVLAELMGVSLMFP-WGFA-----NEITHAKR---183
D 66 KVIVASVALTAALAGSVTSNYM--VLLAIAIPYGLAGVDALNNVVAIHESRIMS 123
QY 184 -FYALFGANISLLASGRATVMSKLRASVSEGD--PMGIS-----RLMAAMTIVSGLV 237
D 124 WILHGMVGASVGYIMGYAL-----SQGGMFMGRRYAILIQVMTLVLSLP 173
QY 238 LMASYMNKNVLTDPFYNEEMOKKKGAKPKNMKDSFLYDRSPY---ILL---290
D 174 LMKRGAIAVESITD-----TASDGAERFGTAEGVSVAERKPLGAVGLAIRGA 225
QY 291 -TLIVAVGICINLIET---WKSQKIQYPMNDYSFPMGNFPMGVSVLIMLFPVG 346
D 226 KEILVMPF--CYCAVESTAGLMASSYMWHSIGIKTA---ASWAS-----LFYGI 272
QY 347 NVIRKF--GMLTGLATVPVWVLLTGIVFALVIFRNQASGLVAMFGTTPMLAVVVAION 405

```

```

D 273 TVGRALSGFLTRFKDPMILGQVIFA-----GILTWVPLPHHLGVVGLVVI 323
QY 406 ILSKSTK-ALPDSKEMAYIPLDQEQVKAKAIDVVAARFGSGGALIQ---GLLVI 461
D 324 GFGCAPYPCVISHTP--AYGEDKSOAIVG---VOMACAV---GSLMPPLFGIIAQ 374
QY 462 CGSIGAMTPYLAVILFIIT 482
D 375 YATISLPWYLLVLLVLMVAM 395

RESULT 40
ID Q9V250 PRELIMINARY; PRT; 976 AA.
AC Q9V250;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Oligosaccharyl transferase.
GN PAB2202.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248283; CAB49148.1;
DR InterPro: IPR003674; I190_trans_STT3.
DR Pfam: PF02516; STT3; 1.
KM Complete proteome.
SQ SEQUENCE 976 AA; 109052 MW; D27A162C72E51A68 CRC64;

Query Match 5.1%; Score 134; DB 17; Length 976;
Best Local Similarity 20.9%; Pred. No. 0.18;
Matches 107; Conservative 79; Mismatches 157; Indels 168; Gaps 29;

QY 59 GAEAIPTKFWLVVPCAIIFLVIY-----KLSNLSQAL--F 95
D 19 GKSILYPLK-RILPLAVIGGIIAYLYRHLTACKRPDPDTFHFETIKLVKEGLPKY 77
QY 96 Y-----AVGTPPLFFALPFTVYPLRDVLPHTPEADRLQAILP--PP--GLGLVIA 142
D 78 YPMAEAPFGSLIGEPGLY--ILPAIFYKVSVGVNEF---QAFIMPPFVFLGVIA 131
QY 143 -----ILRWTFPAFYVLAELMGVSLMFMWGFANEITKIHAKRFA--LFGIGA 192
D 132 VYLLGRKVLNEM-----AGLMAAVILSVSTANFSRTSGNARGDGPFTLFLFSLVA 183
QY 193 NISLSAG--RAIVMS--KLRAVSSEGV---DPMGISRLMAAMTIVSGVLMASVW 244
D 184 MLYTLKENDIKKSLMAVAVFLASISLGAANGSPFLAM-----LIPASQOT 232
QY 245 INKNVLTDPFYNEEMOKKKGAKPKNMKDSFLYDRSPYLLTLVLVAVGICINLI 304
D 233 IALFI-----FGKIKELKKVKEFP-----AYLAIALIGVLT---267
QY 305 EVTWKSQKIQYPMNDYSEFMGNFSP--WTGVSVLIML-FVGNVIRKFGMLTGLVTP 362
D 268 -----PQIAIKIGFI--KFAFEVFLGLVLTIMLYGKGFVNSDKGRFAVVA 314
QY 363 VMVLTGIVFALVIFRNQASGLVAMFGTTPMLAVVVAIONLSSTKVALDSTKEM 422
D 315 VIVLGG---FA-----GAYAVG--PKLRLMGSAVQ---STQ--VQVQDEL 352
QY 423 AYIPLDQEQVKAKAIDVVAARFGSGG---ALIQGGLVTCG-----SIGAMT 469
D 353 A-----KTTLSIDIKLYYGVGNNGVLFSLIPGILLILGLYINALLKSSSSN 400

```

```

Oy      470 PYLAVILLFIITAIWTSATKLKFLAQSAL 500
          | : : : : : : : | | | | :
Db      401 EYMLSLVEFYIMSLYLISLA-VRFLELASVYAI 430

```

Search completed: November 25, 2003, 10:16:11
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: November 25, 2003, 10:20:29 ; Search time 3156 Seconds
(without alignments)
3966.035 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKTEKPFGLRSFLMPH.....AQSLKEQVAQEDSAPASS 515

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09869433/runat_24112003_171459_18132/app_query_faasta_1.711
-DB=EST -QFMT=fastchap -SUFFIX=g2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09869433 -CGN 1.1 2810 -runat_24112003_171459_18132 -NCPU=6 -ICPU=3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST :
1: em_estbda :
2: em_esthum :
3: em_estin :
4: em_estmu :
5: em_estov :
6: em_estpl :
7: em_estro :
8: em_hic :
9: gb_est1 :
10: gb_est2 :
11: gb_hic :
12: gb_est3 :
13: gb_est4 :
14: gb_est5 :
15: em_estfun :
16: em_estom :
17: em_gss_hum :
18: em_gss_inv :
19: em_gss_pln :
20: em_gss_vrt :
21: em_gss_fun :
22: em_gss_mam :
23: em_gss_mus :
24: em_gss_pro :
25: em_gss_rtd :
26: em_gss_phg :
27: em_gss_vr1 :
28: gb_gss1 :

29: gb_gss2 :*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1198	45.6	1866	11 AY107360	AY107360 Zea mays
2	842.5	32.0	966	10 BE574942	BE574942 D06 Triph
3	743.5	28.3	822	14 CB644624	CB644624 OSJNEB050
4	723	27.5	724	14 CB655122	CB655122 OSJNEC08D
5	694.5	26.4	928	14 HS5010	HS5010 HHT42a Scrg
6	684	26.0	645	9 AV926113	AV926113 AV926113
7	668.5	25.4	787	12 B1177811	B1177811 EST518756
8	656	24.9	827	14 CB651454	CB651454 OSJNEB16G
9	629	23.9	576	14 CB604098	CB604098 3529 1 54
10	623.5	23.7	893	14 CD376868	CD376868 PTMMD237
11	618.5	23.5	685	12 BM448586	BM448586 DSA025G04
12	617.5	23.5	705	13 BQ813233	BQ813233 1030034G1
13	609	23.2	757	14 CD423751	CD423751 SAI 1 E12
14	608	23.1	678	9 AJ432835	AJ432835 AJ432835
15	607.5	23.1	744	10 BF460035	BF460035 072A12 Ma
16	605	23.0	600	10 BG097617	BG097617 EST462136
17	582.5	22.1	637	12 BJ482169	BJ482169 BJ482169
18	580	22.1	578	12 BM062387	BM062387 KSO1042A1
19	571	21.7	609	12 BM322252	BM322252 PIC1 2 B0
20	568	21.6	691	10 BG269363	BG269363 10-315T3
21	564	21.4	672	13 BU650245	BU650245 1112085FO
22	555	21.1	567	10 AW931058	AW931058 EST356901
23	553	21.0	664	12 BI405734	BI405734 115G06 Ma
24	550	20.9	520	14 CB381366	CB381366 3529 1 46
25	548	20.8	712	28 BH988329	BH988329 oet186d03
26	547	20.8	715	10 BE459811	BE459811 EST415103
27	545	20.7	641	12 BJ478873	BJ478873 BJ478873
28	541	20.6	510	13 BQ813682	BQ813682 1030038C0
29	537.5	20.6	532	13 BQ469843	BQ469843 HX01A14T
30	536.5	20.4	558	13 BQ790720	BQ790720 E4555 Chi
31	536.5	20.4	567	12 BM954156	BM954156 sam70H06
32	536.5	20.4	700	13 BQ147136	BQ147136 NFO34F08F
33	533.5	20.3	573	12 BJ300874	BJ300874 BJ300874
34	533.5	20.3	611	13 BU044109	BU044109 PP_LBA001
35	533	20.3	602	13 BQ754402	BQ754402 EBCa01 SQ
36	529	20.2	483	9 AV393036	AV393036 AV393036
37	529	20.1	604	9 AV941826	AV941826 AV941826
38	528	20.1	557	9 AW671962	AW671962 LG1 353 G
39	525.5	20.0	563	13 BQ490006	BQ490006 26-E9427-
40	523.5	19.9	519	10 BG454651	BG454651 NF101H03L
41	520.5	19.8	675	10 BE726529	BE726529 894092F04
42	520	19.8	648	9 AW255797	AW255797 ML866 DEP
43	513	19.5	774	28 BH529905	BH529905 BOCGTZ53TF
44	512.5	19.5	766	14 CA783387	CA783387 sat23e03
45	510.5	19.4	568	10 BG463286	BG463286 EMI_48_F1

ALIGNMENTS

RESULT 1
LOCUS AY107360 1866 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PC0098455 mRNA sequence.
ACCESSION AY107360
VERSION AY107360.1 GI:21210438
KEYWORDS
SOURCE HTC
ORGANISM Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1866)

FEATURES	source
1. 1866	/organism="Zea mays"
	/mol_type="mRNA"
	/db_xref="MaizeDB:635637"
	/db_xref="taxon:4577"
	/clone_lib="Maize Mapping Project/Dupont Cornsensus Library"
	/note="this sequence is part of a project of ESR assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT	437 a 446 c 467 g 516 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3.4e-120 Length: 1866
Score:	1198.00 Matches: 240
Percent Similarity:	70.77% Conservative: 82
Best Local Similarity:	52.75% Mismatches: 124
Query Match:	45.55% Indels: 9
DB:	11 Gaps: 3
US-09-869-433-2 (1-515) x AY107360 (1-1866)	
QY	55 A1aPrgIySerIyAlaGluAlaIlePProPheIlySPheTIPLeuValaIProCys 74
Db	5 GCCAAGGCGCAGCGCGCGAGATCATCCCGTTCTCAAGACCTGGGTCAACCTGCCATG 64
QY	75 A1aIleIlePheMetLeuIleTyrAlaIySleuSerAsnIleIleuSerIyGlnAlaIleu 94
Db	65 GCATGCGCGTCAAGTCCCTCTACACCAAGCTCGCGAGTCTCTCCAAAGAGCGCTC 124
QY	95 PheTyrAlaIyAlaGlyThrProPheIleuIlePhePheAlaIleuPheProThValIleTyr 114
Db	125 TTCTAGCGCCTCATCTTCCCGTTCAATCGCCTTCTTGAGGCTTGCGCTAGTGTCTAC 168
QY	115 ProLeuArySPaIleuHisProThrgIuPheAlaAspArgIeuGlnAlaIleuPro 134
Db	185 CCATGGCGCAGCCATCCACCCCACTGGCTCGCGATCGCCTCGCTCGCTCGCGG 244
QY	135 ProGlyIleuIleuIyLeuValaIleAlaIleuAryAsnTTrpThrPheAlaAlaPheTyrVal 154
Db	245 CCCAGCTTCTCGGACCCGCTGGCATCTCCGCGTGTGAAGTTGTGCTCTTACATC 304
QY	155 LeuAlaGluIleuTTrpGlySerValMetLeuSerMetPheTTrpGlyPheAlaAsnIleu 174
Db	305 ATGGCCGAGCTCTGGGCGATGTCGTCATATCCGTCCTTTCTGGGGGTTTGCAATCAG 364
QY	175 ILeThrIyIleHisGluAlaIySArpPheTyrAlaIleuPheGlyIleGlyAlaAsnIle 194
Db	365 ATTAACAAGTGTGAAGAGCCCAAGAGTTCTACCCGCTGTTGGGCTTGGGCGCAATGTG 422
QY	195 SerLeuLeuAlaSerGlyATGAlaIleIleValITPAlaIleTyrSleuAryAlaSerValSer 214

Db	425	GGCCCTCAATCTCTCTGGGCGCAGCGTGAATAATTTCTCAACATGAGGAGAAGATTTGGGT	484
Oy	215	GIuGIyValaApProTfRgIyILSerIeuArgIeuIeuMetAlaMetThrILValaSe	234
Db	485	CTGTGAGAGGATGGTTGGGCAATTTCTGTTGAAGGACATGATGACATATGAGTGGTACT	544
Oy	234	RGILyIeuValIeuMetAlaSerIyTfRTPRILeSnIySaSnValIeuThrAspProAr	254
Db	545	GGGCTCTTCAATCAACCGGTATCTATTTGGGAGGTGAACAAGTTGTTATGATTAAC----	599
Oy	254	gPheTyraSnProGluGluMetGlnIySgIyLyIySgIyValaIyPProIyMeIaSnIyE	274
Db	600	-----TCATCTGTCGCCGTGGTTGAACGGAAAAAGAAAGACCAAGCACAAGCTCAGCAT	652
Oy	274	tLyAspSerPheIeuTyTfLeuAspAygSerProTyTfILeIeuIeuIeuThrLeuVala	294
Db	653	GGGTGAGAGCATGAAAGTTCTGTTGTCTATCTCGGATGTGAGGGAATCTTGCTACATTTGGT	712
Oy	294	ILILaLITyTfGILyILCySILeasIeuIeuILeGILValaThrTfPlySserGILeIyLe	314
Db	713	TGTTCTTATGGAATTAAGCATTAACCTTGTCGAGGTGACATGGAATCAAAATTGAAGC	772
Oy	314	uGILTyProAsnMetAsnAspTyTfYserGILuPheMetGILyAsnPheserPheTfPThGIL	334
Db	773	ACAATTCOCCAAGCCCGAATGAATTTCTTCAATTCATGGCGCATTTCTCAATGCCACCGG	832
Oy	334	yValaIserValIeuIeuMetLeuPheValGILyILyAsnValILeAGILyPheGILyTf	354
Db	833	CATAGCTCAATTTCAATGATGTTG--TTTGGAAAGTGTATCTCCAAAGATTGGGGTG	889
Oy	354	pLeuThrGILyILaIeuValThrProValMeValIeuIeuThrGILyILeValPhePheAl	374
Db	890	GGAGGTGCACACTGCAGTACGCGCTCGAGTGTATCTCTCAAGAGTGGGTTCTTCTC	949
Oy	374	aLeuValILePheArgAsnGILaIserGILyIeuValaIaMetPheGILyThrThrProLe	394
Db	950	ACTGATTTTGTGTTGGTGAACCATTTGACCTTCCTTATGACCAAGTTGGAGATGACCTTT	1009
Oy	394	uMetLeuILaValaIyValGILyILaILeGILaSnILeIeuSerIySserThrIySTyTfAL	414
Db	1010	GCTTGCGCGAGTCTATGTTGGAGCAATGCAACAATTTTCATGAAGATGCACAAATACAG	1065
Oy	414	aLeuPheAspSerThrIySgILuMetAlaTyTfILeProIeuAspGILuGILyValIy	434
Db	1070	TCTGTGTGATCTCTGCACAAGATGGCATACATTTCTTGATGGATGGATGAAAGGTGA	1122
Oy	434	SGILySaILaILaILeAspValaIaIaIaIaArgPheGILySserGILyILaIeuIL	454
Db	1130	AGGTAAAGCGGCTTATGATTTGTGTGCAACCCCTGGGGAATCTGGAGGTCCTGAT	1189
Oy	454	eGILGILyIeuIeuValILeCySgIySerILeGILyALaMetThrProTyTfLeuAlaVa	474
Db	1190	CCACAGCATGATCTCGTCAATTTGGTTCTCTCGCGAATCTGCACACCTTACCTTGGCGG	1245
Oy	474	ILILeIeuIeuPheILeILaILaILeTfPLeuValSerILaThrIySILeIeuIyLeuPh	494
Db	1250	AATATCTCTGTGTGATTTGCTTGCAAGCTGGGTGGTGTAAAGTTCCTCGACCTGCAGTT	1305
Oy	494	eLeuAlaGILserAlaIeuIySgILuGILuIValaILaGILuIu	508
Db	1310	C-----TCTTCCTGGCGAAGCAAGAGCTGGCAGAAAG	1343
RESULT 2			
BE574942			
LOCUS	966 bp	mRNA	linear
DEFINITION	D06 Triphysaria versicolor root-tip, early DMB0-induced transcript		
ACCESSION	BE574942		
VERSION	BE574942.1	GI:12001272	
KEYWORDS	EST.		
SOURCE	Triphysaria versicolor		
ORGANISM	Triphysaria versicolor		

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 966)
Matvienko, M., Torres, M.J. and Yoder, J.I.
Transcriptional responses in the hemiparasitic plant *Triphysaria*
versicolor to host plant signals
Plant Physiol. 127 (1), 272-282 (2001)
21437952
11553755

Contact: Yoder, J.I.
John I. Yoder Research Lab, Dept. of Vegetable Crops
University of California at Davis
137 Amundson Hall, One Shields Drive, Davis, CA 95616, USA
Tel: 530 752 1741
Fax: 530 752 9659
Email: jiyoder@ucdavis.edu
length = 966 bp.

FEATURES
SOURCE
Location/Qualifiers
1..966
/organism="Triphysaria versicolor"
/mol_type="mRNA"
/db_xref="taxon:64093"
/tissue_type="root-tips ~5mm length"
/dev_stage="3-4 weeks growth"
/lab_host="E. coli"
/clone_lib="Triphysaria versicolor root-tip, early
DM60-induced transcript cDNA library"
/note="Vector: PCR2.1 TA Cloning System, Invitrogen,
Carlsbad, CA; PCR-based suppression subtractive
hybridization cDNA library"
BASE COUNT 222 a 186 c 240 g 318 t
ORIGIN

Alignment Scores:
Pred. No.: 1.49e-81 Length: 966
Score: 842.50 Matches: 163
Percent Similarity: 68.09% Conservative: 61
Best Local Similarity: 49.54% Mismatches: 96
Query Match: 32.03% Gaps: 9
DB: 10 Indels: 2

US-09-869-433-2 (1-515) x BE574942 (1-966)

QY 84 LysleuSeranilleuSerlyeGlnalaleuPheTyralaValglyThrProPheleu 103
Db 6 AAGTGGCTAATGTTTGTGCAAGATGCACTTTCTACACTGTTATGTTCCATTCAT 65
QY 104 lIlePhePheAlaleuPheProThValIleTyProleuMaGaPArValleuHleProTh 123
Db 66 GCGTTTGGGGCAATTTGGGTTTGGTTATATCTCTCAGTAATTTTTCACCCCACT 125
QY 124 GluPheAlaAspArgLeuGlnAlaIleuPProFroglyLeuGlyLeuValAlaIle 143
Db 126 GCGCTTCTGATAGCTTTTGAAGCTTGGGTCGAGGTTCTTGGACCACTTGCATTT 185
QY 144 LeuArgAnThrPheAlaAlaPheTyValleuAlaGluLeuTyrglySerValmet 163
Db 186 TTGAGGATTTGGACCTTTGTTGTTCTATGTTATGCTGAGCTTGGGCGAGTGTGTT 245
QY 164 LeuSerleuMetPheTyrglyPheAlaAsnGlnIleThrlyslleHleGlnAlaIleArg 183
Db 246 GTTCTGTTGTTGTTGGGTTTGGGTTTGGCAATCAATCAACCAATCAACCAAGCA 305
QY 184 PheTyValleuPheGlyIleGlyAlaAsnIleSerleuLeuAlaSerGlyArgAlaIle 203
Db 306 TTCTATCTCTTTTGGGCTCGAGCAATGTACACTCATTTCTCGGCTCGAGCTGTG 365
QY 204 ValTrpAlaSerlyLeuArgAlaSerValSerGluGlyValaPProTyrglyIleSer 223
Db 366 AATACTTTTCCAAATGAGCAAAATTTGGCCCGGCTGTGATGTTGGGCGCATCTCG 425

QY 224 LeuArgleuMetAlaMetThrIleValSerGlyLeuValleuMetAlaSerTyTrp 243
Db 426 TTGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 485
QY 244 TrpIleAsnlysaenValleuThAspProArgPheTyrsAnProGluGluMetGlnlys 263
Db 486 TGGGTAAACAAT-----TATGTTCTCTTCGACCCGATGT 521
QY 264 GlylyslsGlyAlaLyProlysmetAsmetlysaSerPheleuTyrlleuAspArg 283
Db 522 ATGAAGAAG 581
QY 284 SerProTyrlleuLeuLeuThleuThleuValIleAlaTyrglyIleCysIleAsnleu 303
Db 582 AAGAAATATATTTAGAGATTTGGCCACTTGTGTGTGTGTGTGTGTGTGTGTGTGT 641
QY 304 lIleGlyValThrTrpIlySerGlnleuLyseGlnleuTyTrProaenMetAsnAspTyser 323
Db 642 GTGAGGTTACATGGAATCAATCAAGCTCAAGCTCAGTTCACCAAGCGGAGAGATATCT 701
QY 324 GluPheMetGlyAsnPheserPheTyrglyValValaSerValleuIleMetleuPhe 343
Db 702 GCTTCATGAGGTGCTTTCATCACTACTACCGGA---ATAGCACTTTCATATATGCTT 758
QY 344 ValGlyGlyAsnValIleArglyPheGlyTrpIleuThrglyAlaLeuValThProVal 363
Db 759 TTGAGTCAATGATGATATTCACCAATACGATGGGGAATGGCAGCAAGATTAACCGC 818
QY 364 MetValleuLeuThrglyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSer 383
Db 819 GTCTGCTAATCACTGGGATTTGGCTTCTCGTGTATCTTTTCGGAGACCTTTGGC 878
QY 384 GlyleuValAlaMetPheGlyThrThProleuMetleuAlaValaIleAlaIle 403
Db 879 CCTGCTCTCGCCGCTTCGGAATGACTCTCTATTAACAGCTGTCTATGTTGGGCGATG 938
QY 404 GluAsnIleLeuSerlySerThrlye 412
Db 939 CAATCAATTTTATGATGAGCGCAAG 965

RESULT 3
CB644624
LOCUS
DEFINITION
OSUNB05020 f OSUNB Oryza sativa (japonica cultivar-group) cDNA
clone OSUNB05020 5', mRNA sequence.
ACCESSION
CB644624
VERSION
CB644624.1 GI:29639615
SOURCE
ORYZA
ORYZA sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 822)
Jandauryarar, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.,
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cga cga gta gta
BACKWARD: gga aac agc tat gac cat g
Plate: 05 row: O column: 20
Seq primer: gta aaa cga cga cga gta
Location/Qualifiers

FEATURES

source

1. 822
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB05020"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)"

BASE COUNT 190 a 166 c 231 g 235 t

ORIGIN

Alignment Scores:

Pred. No.: 8,87e-71 Length: 822
Score: 743.50 Matches: 142
Percent Similarity: 71.22% Conservative: 56
Best Local Similarity: 51.08% Mismatches: 75
Query Match: 28.27% Indels: 5
DB: 14 Gaps: 2

US-09-869-433-2 (1-515) x CB644624 (1-822)

Oy 147 TTTTThPheAlaAlaPheTyrValLeuAlaGluLeuTTPGlySerValMetLeuSerLeu 166
|||
|||
2 TGGAGCTTCTGCTCTTATGTCATGCGAGCTGCGGGAGGCGTCACTCCCGT 61
|||
|||
Oy 167 MetPheTTPGlyPheAlaAsnGluLeuThrLysIleHISGluAlaLysArgPheTyrAla 186
|||
|||
62 CTCCTTGGGATTCGCAATACGATTACTACGTGAAGAGGCCAAGATTACCA 121
|||
|||
Oy 187 LeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValTrrpAla 206
|||
|||
122 CTGTTGGGCTGGGGCAATGTGCACATCTTCTCGGTCGACAGTGAATACATTC 181
|||
|||
Oy 207 SerLysLeuArgAlaSerValSerLuglyValaLpproTrrpGlyIleSerLeuArgLeu 226
|||
|||
182 TCACACATGAGCGCAATTTGGTCCAGGGGTGACGATGGCGATTTCATGAGGGC 241
|||
|||
Oy 227 LeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyrTrrpTrrLeuAsn 246
|||
|||
242 ATGATGACATACGATGCTTGTGCGCTTGTGATTCGCGGATCTACTCGGAGATTAAC 301
|||
|||
Oy 247 LysAsnValLeuThrAspProArgPheTyrAsnProGluGluMetGlnLysGlyLysLys 266
|||
|||
302 AAGTTTGTAAATTGATAA-----TCAGCTGCGCGCATGTCGAGCGGAAG 349
|||
|||
Oy 267 GlyAlaLysProLysMetAsnMetLysAspSerPheLeuTyrLeuAspArgSerProTyr 286
|||
|||
350 AAGGATTAAGCCCAAGACTGATGAGAGAGATTGAAGTGTGTGTCACTCTCGGAT 409
|||
|||
Oy 287 IleLeuLeuLeuThrLeuLeuValIleAlaTrrpGlyIleCysIleAsnLeuIleGluVal 306
|||
|||
410 GTGAGGATCTTGCACCCCTGCTGCTTGTGATTAAGATTAACCTTGTGAGAGTG 469
|||
|||
Oy 307 ThrTrrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTrrSerGluPheMet 326
|||
|||
470 ACATGAGAGTCAAAAGCTGAAGCAATACCAAGCCCAATGATATTCTTCATTCAATG 529
|||
|||
Oy 327 GlyAsnPheSerPheTrrpTrrGlyValValaSerValIleMetLeuPheValGlyLys 346
|||
|||
530 GGGGATTTCTCGACTGCTACTGCGCATACCAATTACAAATGATTTG---TTAGGAGA 586
|||
|||
Oy 347 AsnValIleArgLysPheGlyTrrpLeuThrGlyAlaLeuValThrProValMetValLeu 366
|||
|||
587 GTAATTCCTAGAGAGTTGGGTGGGAGTCGAGCTCAATACATCCCTCCGACGTGGCTT 646
|||
|||
Oy 367 LeuThrGlyIleValaPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuVal 386
|||
|||
647 CTCACGTGAGTGGATTCCTCTCACTGATTTATTGCGTGAAGCATTTGACCTCCTTATG 706
|||
|||

Oy 387 AlaMetPheGlyTrrpTrrProLeuMetLeuAlaValaIleGlyAlaIleGluAsnIle 406
|||
|||
Db 707 GCCACACTTGGAAAGACACCTTGTGCTGCTGCGGATGATGTTGGGCAATGACAACTT 766
|||
|||
Oy 407 LeuSerLysSerThrLysTrrpAlaLeuPheAspSerThrLysGluMetAlaTyr 424
|||
|||
Db 767 TTCAGTAAGAGTGGGAAGTACAGTTGTTTACCCCTGTAAAGAAATGCAATAC 820
|||
|||

RESULT 4
CB655122 724 bp mRNA linear EST 09-APR-2003
LOCUS OSJNEC08D15.F OSJNEC Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEC08D15 5', mRNA sequence.
ACCESSION CB655122.1 GI:29658847
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
1 (bases 1 to 724)
Jantaseuriyarac, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: D column: 15
Seg primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1. 724
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEC08D15"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEC"
/note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2:
XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

BASE COUNT 128 a 231 c 194 g 171 t

ORIGIN

Alignment Scores:

Pred. No.: 1.33e-68 Length: 724
Score: 723.00 Matches: 131
Percent Similarity: 75.43% Conservative: 44
Best Local Similarity: 56.47% Mismatches: 57
Query Match: 27.49% Indels: 0
DB: 14 Gaps: 0

US-09-869-433-2 (1-515) x CB655122 (1-724)

Oy 19 ILeHsthRhISGluLeuLysValLeuProwMetPheLeuMetPhePheCysIleThr 38
|||
|||
Db 21 GTGAGAGCAAGACGCTGGAAGAGATGCGCTGGGCTCATGTTCTTGCATCTC 80
|||
|||
Oy 39 PheAsnTrrpValLeuArgAspThrLysAspThrLeuIleValaIleGlyAlaProGlySer 58
|||
|||

Db 703 TTGGATGAGGATATGAAGGTTAAGGTAAGCAGATTGACGTTGTCTGCA-CCCTT 761

Qy 447 GYLYSERYGLYVALLALEUILLGNGINGLYLEUVALILECYSGYSERIIEGLY 466

Db 762 GGGAAATCTGGAGGCTTTGATCCAGCATTCATCTGTCATTTGATCTCTGG 821

Qy 467 AIAmetThProTyrleuAlaValIleleuPheIleAlaIleIleIleValSer 486

Db 822 AATTGACACCCCTTCTGTTGGATATCTTCTGATTTCTTTCATGCTGGGTCT 881

Qy 487 AIAthThLeuAlaValIleleuPheIleAlaIleIleValSerIleValSer 504

Db 882 GTAAGGCTCTGATCAGACGATT-----TCAACCTGGCAAGCAAGAT 926

RESULT 6
AV926113 645 bp mRNA linear EST 18-JAN-2002
LOCUS AV926113 K. Sato unpublished cDNA library, cv. Haruna Nijo second
DEFINITION leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
clone basd18c14 5', mRNA sequence.

ACCESSION AV926113 GI:18221910

VERSION AV926113.1 GI:18221910

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 645)
Sato, K., Saitoh, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished
Contact: Tadau Shin-ichi
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsbni@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source 1..645
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Haruna Nijo"
/db_xref="taxon:112503"
/clone="basd18c14"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna
Nijo second leaf stage seedling leaves"

BASE COUNT 108 a 202 c 168 g 167 t

ORIGIN

Alignment Scores:
Pred. No.: 2,17e-64 Length: 645
Score: 684.00 Matches: 124
Percent Similarity: 77.57% Conservative: 42
Best Local Similarity: 57.94% Mismatches: 48
Query Match: 26.01% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x AV926113 (1-645)

Qy 25 LysIlySValLeuProMetPheLeuMetPhePheCysIleThrPheAsnTyrThrValLeu 44

Db 2 AAGAAGATCGTCCCTCGGCTTCATGTTCTTCTGATCTCTTCACTACACCATCTCG 61

Qy 45 ArgAspThrIlyAspThrIleuLeuValGlyAlaProGlySerGlyValAlaValIlePro 64

Db 62 AGGGACACCAAGGCTGCTCGTCACCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 121

Qy 65 PheIleIlySValLeuProMetPheLeuMetPhePheCysIleThrPheAsnTyrThrVal 84

Db 122 TTCTCAAGACGTGGGATCAACCTCCCGATCGGCTTTCATGCTTCTACTCCAG 181

Qy 85 LeuSerAnIleLeuSerIlyGlnAlaLeuPheTyrAlaValGlyThrProPheLeu 104

Db 182 CTCGGCAGAGCTCTTCAAGAGGCGCTTTCATACCGGATATCTCCCTTCATCGCC 241

Qy 105 PhePheAlaLeuPheProThrValIleTyrProLeuArgAspValIleuHisProThrGlu 124

Db 242 TTCTTGGCGCTTTCGATACGTGCTTACCTTACCGGAGGAGGAGGAGGAGGAGGAGG 301

Qy 125 PheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeuValAlaIleLeu 144

Db 302 CTCGGCAGACCGCTCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 361

Qy 145 ArgAspThrPhePheAlaPheTyrValLeuAlaGluLeuTyrGlySerValMetLeu 164

Db 362 CGCGCTGAGATTTCTGCTCTTCACTTATAGGGAGCTCTGGGAGGAGGAGGAGGAGG 421

Qy 165 SerLeuMetPheTyrGlyPheAlaAsnGluIleThrIlySileGluAlaIlyAspPhe 184

Db 422 TCCGCTCTGTTTGGGATTTGCCATCAGATTACAGAGGTTGAAGAGGCTTAAAGATT 481

Qy 185 TyrAlaLeuPheGlyIleGlyValAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleVal 204

Db 482 TACCACATGTTGGGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541

Qy 205 TrpAlaSerIlySValLeuAlaSerValSerGluGlyValAspProTyrGlyIleSerLeu 224

Db 542 TACTTCTAAACATAGGAGCAATTTGGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGG 601

Qy 225 ArgLeuMetAlaMetThrIleValSerGlyLeuValLeu 238

Db 602 AAGGCATGATGAGCATGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643

RESULT 7
B1177811 787 bp mRNA linear EST 07-MAR-2003
LOCUS B1177811
DEFINITION EST518756 cSTE Solanum tuberosum cDNA clone cSTE11L7 5' sequence,
mRNA sequence.

ACCESSION B1177811 GI:14643622

VERSION B1177811

KEYWORDS EST.

SOURCE Solanum tuberosum (potato)

ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 787)
van der Hoeven, R., Bezerides, J., Bachem, C., Visser, R., Cho, J.,
Chieningo, A., Bougri, O., Buell, C.R., Ronning, C., Tanksley, S. and
Baker, B.
Generation of ESTs from in vitro grown microtubers
Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr. Rockville, MD 20850, USA
Email: potato-array@igr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
Location/Qualifiers

FEATURES
source 1..787
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTE11L7"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/clone_lib="cSTE"

Db 191 TGGGGTCGGGCTCGCCCAAGTTCCTGGGG-GTGGAGAGAGAGCGCTGAGAGATCTGTG 249
 Qy 29 ProMetPheLeuMetPhePheCysIleThrPheAsnTrpValLeuArgAspThrIys 48
 Db 250 CGCGTGGGGCTCATCTTCTTCTGCACTCTTCTTCACTACACCATCTCGGGGACGCCAG 309
 Qy 49 AspThrLeuIleValGlyValAspSerGlyValGluValIleProPheIleLeuSche 68
 Db 310 GACGCTGCTGCTGCTACCGCCCAAGGCGACAGCGGGAGATCATCTCCGTTCTCTCAAG 369
 Qy 69 TrpLeuValValProCysAlaIleIlePheMetLeuIleTyAlaLysLeuSerAsnIle 88
 Db 370 TGGGTCAACCTCCCATCGCCATCGGGTTCACTGCTCTACACCAAGCTCGCCGACGCTG 429
 Qy 89 LeuSerIysGlnAlaLeuPheTyAlaValGlyThrProPheLeuIlePhePheAlaLeu 108
 Db 430 CTCTCCAGAGAGCGCTCTTCTACACCGCTCATCTTCCCTCATGCGCTTCTCGCGCC 489
 Qy 109 PheProThrValIleTyProLeuArgAspValLeuHisProThrGluPheAlaAspArg 128
 Db 490 TTGGCTACCTGCTTACCCCATGCGGACGCGCATCCACCGCGCGCTCGCCGACGCGC 549
 Qy 129 LeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThr 148
 Db 550 CTCTCGCGCGCGCTGGGCGCCGCTTCTCGGGCGCCGCTCGCATCTCGCGCTCGAGC 609
 Qy 149 PheAlaAlaPheTyValLeuAlaGluLeuTrpGlySerValMetLeuSerLeuMetChe 168
 Db 610 TTCTGCTCTTCTTATGATGATGCGCGACGCTGGGGCAGCGTGTATCTCGCTCTTC 669
 Qy 169 TrpGlyPheAlaAsnGluIleThrIleHisGlnAlaLysArgPheTyAlaLeuPhe 188
 Db 670 TGGGATTCGCGCAATCACTTACTACGCTGTAAGAGCGCAAGATTTTACCACTGTC 729
 Qy 189 GlyIleGlyAlaAsnIleSerLeuLeuAlaSerGlyAlaGlnAlaIleValTrpAlaSerIys 208
 Db 730 GGGCTTGGGGCAATGTCGACACTCATCTTCTCGGTCGACAGTGAATACTTCTCAAC 789
 Qy 209 LeuArgAlaSerValSerGluGlyValAspProTrp 220
 Db 790 ATGAGGCGAATTTGGTTCAGGGGTGACGCGATG 825
 RESULT 9
 LOCUS CB604098 576 bp mRNA linear EST 16-MAY-2003
 DEFINITION 3529_154_1_F08_Y_1 3529 - 2 mm ear tissue from Schmidt and Hake
 ACCESSION CB604098
 VERSION CB604098.1 GI:29543718
 KEYWORDS EST.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 576)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 JOURNAL University
 COMMENT Unpublished
 CONTACT: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 3529_154_1_row_F column: 08.
 FEATURES
 source
 1..576
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultiVar="B73"

/db_xref="taxon:4577"
 /tissue_type="ear"
 /dev_stage="2 mm"
 /lab_host="E. coli XL0R"
 /clone_id="3529 - 2 mm ear tissue from Schmidt and Hake
 lab8"
 /note="Organ: ear; Vector: PAD-GAL4-2.1; Site 1: EcorI;
 Site 2: XhoI; RNA isolated by Hake lab. 1 million pfu
 amplified. Ampicillin is the selection marker."
 BASE COUNT 101 a 199 c 138 g 138 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,03e-58 Length: 576
 Score: 629.00 Matches: 114
 Percent Similarity: 77.89% Conservative: 34
 Best Local Similarity: 60.00% Mismatches: 42
 Query Match: 23.92% Indels: 0
 DB: 14 Gaps: 0
 US-09-869-433-2 (1-515) x CB604098 (1-576)
 Qy 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPheCysIleThrPheAsn 40
 Db 2 ACAGAGCAATCAAGAGATGTCGCCCTGGGGCTCATGTTCTTCTGATCTCTTCAAC 61
 Qy 41 TyrThrValLeuAlaGspThrIlyAspThrIleuIleValGlyAlaProGlySerGlyAla 60
 Db 62 TACACCATCTCCGCGGATACCAAGAGCGTGTCTGTCTACCGCCAAAGGCGAGCGCC 121
 Qy 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 Db 122 GAGATCATCCGCTTCTCTCAAGACTGGGTGCACTGCCATGAGAGTGGCTTATGCTC 181
 Qy 81 IleTyAlaLysLeuSerAsnIleLeuSerIysGlnAlaLeuPheTyAlaValGlyThr 100
 Db 182 CTCTACACCAAGCTCGCGCATGTCTCTCCAAAGAGCGGCTTCTTACGCGCTCATCTTC 241
 Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyProLeuArgAspValLeu 120
 Db 242 CGGTTCATCGCTTCTTCTGGCGCTTGCCTTACGCTTACCCCAAGCGGACGCCATC 301
 Qy 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 Db 302 CACCCCACTGGCTCGCGCATGCGCTCTCGCTCGCGCCCACTTCTCGAGACC 361
 Qy 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyValLeuAlaGluLeuTrpGly 160
 Db 362 GTTGCCATCTCCCGCGCTGGAGCTTGCCTTCTTACGTCATGCGCGAGCTCTGGGCGC 421
 Qy 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrIleHisGlu 180
 Db 422 AGTGTCTCATATCCGCTTCTTCTGGGGTTGCAATCATGATCTTACAGTTGAAGAG 481
 Qy 181 AlAlaLysArgPheTyAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 Db 482 GCCAAAGATTCTACCGCGCTTGGGCTTGGGCGCATGCGCCCTCATCTTCTCGG 541
 Qy 201 ArgAlaIleValTrpAlaSerIysLeuArg 210
 Db 542 CGCACGGTGAATATTTCTCAACATGAGG 571
 RESULT 10
 LOCUS CD376868 893 bp mRNA linear EST 31-MAY-2003
 DEFINITION PTMM02037 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum
 tricornutum cDNA 5', mRNA sequence.
 ACCESSION CD376868
 VERSION CD376868.1 GI:31252482
 KEYWORDS EST.
 SOURCE Phaeodactylum tricornutum
 ORGANISM Phaeodactylum tricornutum
 Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
1 (bases 1 to 893)	Bacillariophyceae; Naviculales; Phaeodactylaceae; Phaeodactylum					
1	Scala, S., Carale, N., Falcicatore, A., Chiavano, M.L. and Bowler, C.	Genome properties of the diatom Phaeodactylum tricornutum	Plant Physiol.	129 (3), 993-1002 (2002)	12111123	
12111555						
12114555	Contract: Bowler C	Laboratory of Molecular Plant Biology				
	Stazione Zoologica 'Anton Dohrn					
	Villa Comunale, I-80121, Napoli, Italy					
	Tel: 39 081 583 3268/3211					
	Fax: 39 081 764 1355					
	Email: chris@alpha.szn.it					
	Diatom EST Database (http://avesthagen.sznbowl.com)					
	Seq primer: T3 backward					
	POLYA=Yes.					
FEATURES						
source	Location/Qualifiers					
	1..893					
	/organism="Phaeodactylum tricornutum"					
	/mol_type="mrna"					
	/db_xref="taxon:2850"					
	/cell_line="CCMP632"					
	/clone_lib="Phaeodactylum tricornutum Uni-Zap XR"					
	/note="Vector: Uni-Zap XR vector; Site_1: Eco RI; Site_2: Xho I"					
BASE COUNT	196 a 220 c 238 g 227 t 12 others					
ORIGIN						
Alignment Scores:						
Pred. No.:	1,496-57	Length:	893			
Score:	623.50	Matches:	129			
Percent Similarity:	78.37%	Conservative:	34			
Best Local Similarity:	53.37%	Mismatches:	56			
Query Match:	23.71%	Indels:	2			
DB:	14	Gaps:	1			
US-09-869-433-2 (1-515) x CD376868 (1-893)						
Qy	261 MetGlnTyrGlyLeuYsgLysGlyAlaValProLysMetAsnMetLysAspSerPheLeuYr	280				
Db	8 CTAAAGAGACTATCTTCGACAAAAGAAAGCCCAATATCTATAAGATGGCCCAAGTTT	67				
Qy	261 LeuAspArgSerProTyrTrileuLeuLeuThreuLeuValIleAlaTyrGlyIleCys	300				
Db	68 CTTTCAGACGCCCTTACATTCGCAGATCTAGCTTTGTGTGTAATTCTGACGCAGATGTC	127				
Qy	301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn	320				
Db	128 ATCAACATTGTGCAATCGAATCGAGTGAAGCAACACTCAAGAGGCTTTCCGATCCGAC	187				
Qy	321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValIleIle	340				
Db	188 TCGTACTCCGCCCTTATGAGGTAACTTCTCTCGACGTACCGG--GGCGCACTTGATT	244				
Qy	341 MetLeuPheValGlyValAsnValIleArgLysPheGlyTyrPheunThrGlyAlaLeuVal	360				
Db	245 ATGATGTTCTCGCGCGCATTCATCTTTCAGAAATTGGTTGGAGAACAGCTGGCGTCTC	304				
Qy	361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn	380				
Db	305 ACTCCAAACGATATCGGTGTGACGGGCTGGGTTCTTCCGCCCAACAGCTTTTCGGCC	364				
Qy	381 GlnAlaSerGlyLeuValAlaMetPheGlyTyrThrProMetLeuAlaValVal	400				
Db	365 GCCATGTCCTCCGGTCCCGCTCTTTTGGGACAAACCCGCTTATGTTGGGCTTGCTC	424				
Qy	401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys	420				
Db	425 GGAGCGCGCCAGAAATTTCTGACCAAGAGCTCCAAATACCTTTGTTGATCCGTGAAA	484				
Qy	421 GluMetAlaTyrTrileProLeuAspGlnGlnLysValLysGlyLysAlaIleAsp	440				

Db	485	GAATGGCGCTACATCCCTTTTGATCAGAAATCCAAACCAAAAGCAGCGCCATGAT	544
Qy	441	ValValaAaAaAGpNhegLyVlySserGly-GLYAlaLeuIleGlnGlnGlyLeuLeuA	460
Db	545	GTTCCTGGCAATCCCTTGGGAAAGAGGTGGGGCGCTGATTCAAACAGGTTCTTATCTT	604
Qy	460	lIleCgylSerIleGlyAlaMetThrProTyLeuAlaValIleLeuLeuPheIleI	480
Db	605	TGAGATCGGATGTTTAAAGCCCGCGCTACAGCCGATTTAGCCGTGATCTTTGCTGGTCTGGT	664
Qy	480	e 480	
Db	665	A 665	
RESULT 11			
LOCUS	BM448586	685 bp	mRNA linear EST 01-APR-2002
DEFINITION	DSAO25G04.59793	An expressed sequence tag database for the halotolerant green alga, Dunaliella salina	Dunaliella salina cDNA
ACCESSION	BM448586		
VERSION	BM448586.1	GI:19854158	
KEYWORDS	EST.		
SOURCE	Dunaliella salina		
ORGANISM	Dunaliella salina		
REFERENCE	Eukaryota, Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Dunaliellaceae; Dunaliella.		
AUTHORS	1 (bases 1 to 685)		
TITLE	Cushman,J.C.		
JOURNAL	An expressed sequence tag database for the halotolerant green alga, Dunaliella salina		
COMMENT	Unpublished		
	Contact: Cushman JC		
	Department of Biochemistry		
	University of Nevada		
	MS200, Reno, NV 89557-0014, USA		
	Tel: 775-784-1918		
	Fax: 775-784-1650		
	Email: jcushman@unr.edu		
	PCR Primers		
	FORWARD: T3 20mer		
	BACKWARD: T7 21mer		
	Plate: 025	row: G	column: 04
	Seq primer: T3 20mer		
	High quality sequence stop: 685.		
FEATURES			
source	1..685		
	/organism="Dunaliella salina"		
	/mol_type="mRNA"		
	/db_xref="taxon:3046"		
	/clone="DSAO25G04"		
	/tissue_type="Cells, which was adapted in 2.5M NaCl via a incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl, were exposed to 3.4 M NaCl for 5 hours"		
	/cell_type="Green"		
	/clone_id="An expressed sequence tag database for the halotolerant green alga, Dunaliella salina"		
	/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapR vector and cDNA synthesis kit."		
BASE COUNT	123 a	214 c	184 g 164 t
ORIGIN			
Alignment Scores:			
Pred. No.:	3,67e-57	Length:	685
Score:	618..50	Matches:	123
Percent Similarity:	73.81%	Conservative:	32
Best Local Similarity:	58.57%	Mismatches:	54
Query Match:	23.52%	Indels:	1
DB:	12	Gaps:	1

Db 596 ATGATGATCTGTCCTCCGTCGATCTTCAAGACGATTCGGCTGGNNGTGTGCTGCCCTGAT 655

QY 360 1ThProValMetValLeuLeuThGlyLeValahepheaLeuVal 376

Db 656 CACCCCAATGATGCTGATGATCAGGCTGCTGCTTTCACCCCTGGTG 704

RESULT 13

CD423751 757 bp mRNA linear EST 02-JUN-2003

LOCUS SAl_1.E12_A002 3', mRNA sequence.

DEFINITION CDNA clone SAl_1.E12_A002 3', mRNA sequence.

ACCESSION CD423751

VERSION CD423751.1 GI:31330014

KEYWORDS EST.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 757)

CDomnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.

An EST database from Sorghum: salicylic acid-treated seedlings unpublished

TITLE Contact: Cordomnier-Pratt MM

JOURNAL Laboratory for Genomics and Bioinformatics

COMMENT The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-1271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmprratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug3 (CGACCTGACGCTCGACACAC)

POLYA=yes.

FEATURES

source

location/Qualifiers

1..757

/organism="Sorghum bicolor"

/mol_type="mRNA"

/cultivar="BTx623"

/db_xref="taxon:4558"

/clone="SAl_1.E12_A002"

/lab_host="DH10B-Tl phage-resistant E. coli"

/clone_lib="Salicylic acid-treated seedlings"

/note="Vector: pME185-FLJ; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 1 mM salicylic acid (SA). Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME185-FLJ vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 191 a 155 c 191 g 220 t

ORIGIN

Alignment Scores:

Pred. No.: 4.65e-56 Length: 757

Score: 609.00 Matches: 127

Percent Similarity: 69.28% Conservative: 40

Best Local Similarity: 52.70% Mismatches: 70

Query Match: 23.16% Indels: 4

DB: 14 Gaps: 2

US-09-869-433-2 (1-515) x CD423751 (1-757)

QY 265 LysIysGlyAlaLysProLysMetAspMetLysAspSerPheLeuTyLLeuAspArgSer 284

Db 46 AAAAAGAGGACAAAGCCAAAGCTCAGATGGAGACATGAAGTTCTGTTCACCT 105

QY 285 ProTyrIleLeuLeuThLeuThLeuValIleAlaTyrGlyIleCysIleAsnLeuIle 304

Db 106 CGGATGTGAGGAGATCTTCCACATGTCCTGTGCTTATGAAATTAAGCATTAACCTGTG 165

QY 305 GluValThrTyrPheSerGlnLeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGlu 324

Db 166 GAGGATCATGGAATCAAAATTTGAAGACAGATCCCAAGCCCAAGAAATTAATCTTCA 225

QY 325 PheMetGlyAsnPheSerPheTyrThrGlyValValSerValLeuIleMetLeuPheVal 344

Db 226 TCACATGGGCGATTTCTCAACTGCCACCGCATATGACTTTCATCATATATGTTG--TTA 282

QY 345 GlyGlyAsnValIleArgLysPheGlyTyrPheThrGlyAlaLeuValThrProValMet 364

Db 283 GGGAGAGTAATCTCAGAAAGTTGGGGAGGAGTTGCAGCTACATCACCTCAGTG 342

QY 365 ValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGly 384

Db 343 TTACTTCTCACAGAGTTGGGTTCTTCTCACTGATTTTGTGTGAGCCATTCATCTCT 402

QY 385 LeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValGlyAlaIleGln 404

Db 403 CTTATGACCAAGTTGGGATGACACCTTGGCTGGGAGAGCTCATGTTGGAGCAATGAG 462

QY 405 AsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLysGlnMetAlaTyr 424

Db 463 AACATTTTCAGTAAGAGTCGCAAGAGTACGCTGTGTTATCTCTGCAAGAGATGCAATC 522

QY 425 IleProLeuAspGlnGlnGlnValLysGlyLysAlaAlaIleAspValValAlaAla 444

Db 523 ATTCCTTTCATGATGAGATATGAAGTTAAAGCTTAAGCATTGACCTTGTGCAAC 582

QY 445 ArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuLeuValIleCysGlySer 464

Db 583 CCCTGGGGAATCTGGAGGCTTGTATCCAGCAGATTCATGATCTTCATTTGATGATCT 642

QY 465 IleGlyAlaMetCThrProTyrLeuAlaValIleLeuLeuPheIleIleAlaIleTyrLeu 484

Db 643 CTTCGGAATTCGACACCTTCTGTCGTAATTAATCTTGGATTTGTTCTTCATGGCTG 702

QY 485 ValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeuLysGlnGln 504

Db 703 GGTGCTGAAGTCCCTGATCAAGTT-----TCAACCTGGCAAGCAAGAT 753

QY 505 Val 505

Db 754 CTC 756

RESULT 14

LOCUS AJ432835 678 bp mRNA linear EST 15-MAR-2002

DEFINITION AJ432835 S00007 Hordeum vulgare cDNA clone S0000700012A06F2, mRNA sequence.

ACCESSION AJ432835

VERSION AJ432835.1 GI:19521287

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

1 (bases 1 to 678)

Triticaceae; Hordeum.

AUTHORS Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.

TITLE Barley EST's

JOURNAL Unpublished

COMMENT Contact: Schulman AH

Institute of Biotechnology

University of Helsinki
P.O. Box 56 (Vikinkaari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES

source Location/Qualifiers

1..678
/organism="Hordeum vulgare"
/mol_type="rRNA"
/db_xref="taxon:4513"
/clone="S0000700012A06F2"
/dev_stage="Shoot"
/clone_1ib="S00007"
/note="2,-3,-4-days after germination"
BASE COUNT 168 a 139 c 166 g 205 t
ORIGIN

Alignment Scores:

Pred. No.: 5,15e-56 Length: 678
Score: 608.00 Matches: 116
Percent Similarity: 72.25% Conservative: 48
Best Local Similarity: 51.10% Mismatches: 59
Query Match: 23.12% Indels: 4
DB: 9 Gaps: 3

US-09-869-433-2 (1-515) x AJ432835 (1-678)

142 AAlaIleleuAaGAntTPhrPheAlaAlaPheTtYValleuAlaGluLeuTtPglYSer 161
142 AAlaIleleuAaGAntTPhrPheAlaAlaPheTtYValleuAlaGluLeuTtPglYSer 161
9 GGTATCCCTCAGATTGGAGCTTCTGCTTATTCATATGTCATGCGAGCTAAGGGGCGAGC 68
162 ValMetLeuSerLeuMetPheTtPglYPheAlaAaGluLeuTtPglYSer 181
69 GTCGTTGCTCTGCTCTCTCTCTGCGGTTTGGCAACCAATCACCAGCTTATGAGA 128
182 LysAaGpHeTtYrAlaLeuPheGlyYleGlyAlaAaGluLeuTtPglYSer 201
129 AAAGAGTCTACCTCTGTTGGCCCTTGGGCAATATGACCCTTATCTTTCCGCGGCT 188
202 AlaIleValtPAlaSerYtLeuAaGluLeuTtPglYSer 221
189 ACTGGAAGATTTCTCAATTCGGAAGACAAAGGCTCGACGAGTGTATGTTGGAG 248
222 IleSerLeuAaGluLeuMetAlaMetThrIleValSerGlyLeuValleuMetAlaSer 241
249 GATCTTTGAAGAAGATATGCGCATAGTACTTCTTGGCTTGTACTTCTTCATT 308
242 TtTtPtlPtlAaAaLysAaAaValleuThraSpProAaGpHeTtYrAaSpProGluLumet 261
309 TATTGGGAGTGAACAAGCTTGTGTAATGACCTTCTCTT-----CCAAATCTGAT 362
262 GlnLysGlyLysLysGlyAlaLysProLysMetAaMetLysAaSpSerPheLeuTtPglY 281
363 CACAAGAAAGAAAAAAGAAC--AAACCTAAGCTTACATGAAGAAGATGAAGATTCTG 419
282 AaPaGpSerProTtYrIleLeuLeuLeuThraLysValleuAlaIleAlaTtGlyLeuCysIle 301
420 ATCTCTTCAAAATATGTAAGGACCTTGTACTTACTTACTTGTGCAATGCAATGATC 479
302 AaAaLeuIleGluValtThrTtPtlPtlAaAaLysAaAaValleuThraSpProAaGpHeTtYr 321
480 AATCTTGGGAAGTCAATGAAATGCAAGCTCAAGGACAGAGTCCCAAGTCCCAAGCAG 539
322 TtYtSerGluPheMetGlyAaAaPheSerPheTtPtlPtlYValleuValleuMet 341
540 TACTCATCTTCATGAGTATTCCTTACTGCAACCTGAATTGCAACTTTCACGATGATG 599
342 LeuPheValtGlyLysAaValleuThraLysPheGlyTtPtlPtlYValleuValtThr 361
600 CTG---TTAGTTCGAGTATATTCACAAAGATTGCTGGGCTGAGCCGCAATGATCAC 656
362 ProValMetValleuLeuThr 368
657 CCCACAGTTTGTGTGTGACT 677

RESULT 15

BR460035

LOCUS

072A12 Mature tuber lambda ZAP Solanum tuberosum cDNA 5, similar to
PLASTIDIC ATP/ADP-TRANSPORTER gl|7441708|p... sp|O24381, mRNA

DEFINITION

BR460035 744 bp mRNA linear EST 11-SEP-2002
Sequence.

ACCESSION

BR460035

VERSION

BR460035.1

KEYWORDS

EST

SOURCE

Solanum tuberosum (potato)

ORGANISM

Solanum tuberosum

REFERENCE

1 (bases 1 to 744)

Crookshanks, M., Emmersen, J., Welinder, K.G. and Nielsen, K.L.

The potato tuber transcriptome: analysis of 6077 expressed sequence

tags

FERS Lett. 506 (2), 123-126 (2001)

JOURNAL 21475600

MEDLINE 11591384

PUBMED 11591384

COMMENT

Contact: Karen G. Welinder

Institut for bioteknologi

Aalborg Universitet

Sohnsgaardsholmsvej 49, 9000 Aalborg, Denmark

Tel.: +45 96358467

Fax: +45 98141808

Email: kgw@bio.auc.dk

High quality sequence stop: 744

POLYA=NO.

location/Qualifiers

1..744

/organism="Solanum tuberosum"

/mol_type="rRNA"

/cultivar="Field grown Kuras"

/db_xref="taxon:4113"

/tissue_type="Tuber"

/clone_1ib="Mature tuber lambda ZAP"

/note="Vector: Lambda ZAP"

BASE COUNT 189 a 149 c 178 g 228 t

ORIGIN

FEATURES

source

1..744

/organism="Solanum tuberosum"

/mol_type="rRNA"

/cultivar="Field grown Kuras"

/db_xref="taxon:4113"

/tissue_type="Tuber"

/clone_1ib="Mature tuber lambda ZAP"

/note="Vector: Lambda ZAP"

BASE COUNT 189 a 149 c 178 g 228 t

ORIGIN

Alignment Scores:

Pred. No.: 6,64e-56 Length: 744

Score: 607.50 Matches: 124

Percent Similarity: 66.80% Conservative: 47

Best Local Similarity: 48.44% Mismatches: 76

Query Match: 23.10% Indels: 10

DB: 10 Gaps: 2

US-09-869-433-2 (1-515) x BR460035 (1-744)

175 IleThyLysIleHisGluAlaLysAaGpHeTtYrAlaLeuPheGlyYleGlyAlaAaGlu 194
3 ATCAAGACTGTCAGATGAGGCTAAGAGATTCATCTTGTGTTGGACTTGGAGCAAGTGT 62
195 SerLeuLeuAlaSerGlyAaGluAlaIleValtPtlAaSerLysAaGluAlaSerValSer 214
63 GCTCTTATTTCTCTGTCGACAGTGAAGATCTTCTTACTTGAAGAAGCTTTTGAAGT 122
215 GluGlyValaAaPtlPtlYleSerLeuAaGluLeuMetAlaMetThrIleValSer 234
123 CTGAGAGTGAATGTTGGGCAATCTTCTTGAAGAAGATGATGATGATGATGATG 182
235 GlyLeuValleuMetAlaSerTtTtPtlPtlAaAaLysAaAaValleuThraSpProAaGp 254
183 GGTGGGGAATCTGTTCTTACTGAGTGGTGAATGAATGTTGCTCTCCAACTCGT 242
255 PheTtYrAaSpProGluLumetGlnLysGlyLysLysGlyAlaLysProLysMetAaMet 274
243 -----AGCAAGAAAGAAAGAGTAAACCTTAACATGACCA 278
275 LysAaSpSerPheLeuTtPtlPtlAaAaGpHeTtYrIleLeuLeuThraLeuVal 294


```
Db      279  ATGGAGAGCTTGAAGTCTTGATCTCTTCAAAATATATCAGGATCTTGCCACATGGTT 338
      295  ILeaLaTYrGIyIleCyIleasnleuileguValThrTpySerGlnleuLeu 314
      339  GTACCATATGGCATTAAGTATCAACCTTGTAAGTTACATGGAAGTCAAGCTCAAGCT 398
Qy      315  GlnTYrProAsnMetAsnAspTYrSerGlnPheMetGlyAsnPheserPheTyrThrGly 334
      399  CAGTTCACAGCCCAATGAATACTCTCATTCATGGGAGACTTCTCACTGCTACTGGA 458
Qy      335  ValValSerValleuIleMetleuPheValGlyIleAsnValIleargLysPheGlyTyr 354
      459  ---ATAGCAACTTCATATATGATGTTGTTAAGTCAATGATTTTCGACAGATGGGTGG 515
Qy      355  LeuThrGIyAlaLeuValThrProValMetValleuLeuThrGIyIleValPhePheAla 374
      516  GGAGCAGCAGCCCAAGATACACTACAGTCTTGCTCTTACCGGAGTTGTTCTTCTCC 575
Qy      375  LeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeu 394
      576  CTGCTTTTGTGGTGGCCCTCTAGCACCTACTCTTGCGAAGTTGGATGACTCTCTT 635
Qy      395  MetLeuAlaValIleValGIyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAla 414
      636  CTAGAGCTGTCTATGTGGTGCAATGCAGAACATTTC-AGTAAGAGTGCAAGATATAGT 694
Qy      415  LeuPheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGlu 430
      695  TTGTTTGACCCCTCGAAGAAATGGGCTACATTCCTTTGGATGAGGAC 742
Db
```

Search completed: November 25, 2003, 12:56:46
Job time : 3169 secs

THIS PAGE BLANK (USP 10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 25, 2003, 12:10:00 ; Search time 477 Seconds
(without alignments)
3557.160 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 2630
Sequence: 1 MTKTEKPFGLRSLFPIH.....AQSALKEQVAGEDSAPASS 515

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 219069 seqs, 1647345023 residues
Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.epool/US09869433/runat_24112003_171501_18234/app_query.fasta_1.711
-DB=Published Applications NA -OPMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=plc -NORM=ext -HEAPSIZE=500 -MINTLEN=0
-MAXLEN=200000000 -USER=US09869433 @CGN_1_1_347 @runat_24112003_171501_18234
-NCPU=6 -ICPU=3 -NO MAP -LANG=QUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2630	100.0	1637	9	US-09-892-851-1 Sequence 1, Appl1

2	1333	50.7	1875	10	US-09-938-842A-694	Sequence 694, App
3	256	9.7	246	9	US-09-294-093B-4236	Sequence 4236, App
4	242	9.2	284	9	US-09-294-093B-5730	Sequence 5730, App
5	227.5	8.7	309	9	US-09-294-093B-1312	Sequence 1312, App
6	162	6.2	1614	10	US-09-738-626-117	Sequence 117, App
7	162	6.2	3309400	10	US-09-738-626-117	Sequence 1, Appl1
8	138.5	5.3	1389	9	US-09-815-242-6064	Sequence 6064, App
9	138	4.84	484	11	US-09-770-961-833	Sequence 833, App
10	133	5.1	1830121	14	US-10-329-960-1	Sequence 1, Appl1
11	132.5	5.0	1392	9	US-09-815-242-6047	Sequence 6047, App
12	122.5	4.7	68750	13	US-10-014-717-1	Sequence 1, Appl1
13	120.5	4.6	1188	9	US-09-815-242-9794	Sequence 9794, App
14	120.5	4.6	1278	9	US-09-815-242-8073	Sequence 8073, App
15	120.5	4.6	76804	12	US-10-085-959-109	Sequence 109, App
16	120	4.6	1503	10	US-09-938-842A-1892	Sequence 1892, App
17	119.5	4.5	1434	9	US-09-815-242-9742	Sequence 9742, App
18	119.5	4.5	5132	8	US-08-781-986A-163	Sequence 163, App
19	119	4.5	1789	9	US-09-815-242-4189	Sequence 4189, App
20	119	4.5	1815	9	US-09-815-242-8427	Sequence 8427, App
21	119	4.5	12445	10	US-09-070-927A-242	Sequence 242, App
22	118	4.5	1401	9	US-09-815-242-9733	Sequence 9733, App
23	117.5	4.5	1290	9	US-09-815-242-4623	Sequence 4623, App
24	117.5	4.5	3642	8	US-08-781-986A-328	Sequence 328, App
25	117.5	4.5	77536	12	US-09-940-316B-1	Sequence 1, Appl1
26	116.5	4.4	1684	10	US-09-881-752A-135	Sequence 135, App
27	116	4.4	1467	14	US-10-156-761-4216	Sequence 4216, App
28	116	4.4	25801	12	US-10-181-319-13	Sequence 13, Appl1
29	116	4.4	9025608	14	US-10-156-761-1	Sequence 1, Appl1
30	115.5	4.4	1230	10	US-09-738-626-361	Sequence 361, App
31	115	4.4	1176	12	US-10-181-319-3	Sequence 3, Appl1
32	115	4.4	2124	10	US-09-738-626-1524	Sequence 1524, App
33	115	4.4	3309400	10	US-09-738-626-1	Sequence 1, Appl1
34	114	4.3	12860	10	US-09-070-927A-144	Sequence 144, App
35	113.5	4.3	1830121	14	US-10-329-960-1	Sequence 1, Appl1
36	113	4.3	3867	12	US-10-171-319-10	Sequence 10, Appl1
37	113	4.3	5668	9	US-09-759-143-777	Sequence 777, App
38	113	4.3	5668	9	US-09-780-669-777	Sequence 777, App
39	113	4.3	5668	9	US-09-892-827-777	Sequence 777, App
40	113	4.3	5668	10	US-09-895-793-777	Sequence 777, App
41	113	4.3	5668	10	US-09-895-814-777	Sequence 777, App
42	113	4.3	5668	12	US-10-144-678A-777	Sequence 777, App
43	113	4.3	5668	12	US-10-294-025-777	Sequence 777, App
44	113	4.3	5668	13	US-10-012-896-777	Sequence 777, App
45	113	4.3	5668	14	US-10-205-823-420	Sequence 420, App

ALIGNMENTS

RESULT 1
US-09-892-851-1
Sequence 1, Application US/09892851
Patent No. US20020081682A1
GENERAL INFORMATION:
APPLICANT: Aventis Pasteur Limited
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
FILE REFERENCE: 77813-6
CURRENT APPLICATION NUMBER: US/09/892,851
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/114,060
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: US 60/123,967
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/141,271
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1637
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (51)..(1595)

US-09-892-851-1

Alignment Scores:
 Pred. No.: 9,04e-294 Length: 1637
 Score: 2630.00 Matches: 515
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-869-433-2 (1-515) x US-09-892-851-1 (1-1637)

QY 1 MetThrLysThrGluGluLysProPheGlyLysLeuArgSerPheLeuTrpProIleHis 20
 DB 51 ATGACAAAACCGAAGAAAACCTTTTGGAAAATTGCGCTCTTTCTTGCGCGCATACAT 110
 QY 21 ThrHisGluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 DB 111 ACTCAGAGCTAAGAAAGTTCTGCGCAATGTTCTTAATGTTCTTCTGATTAATTAAAC 170
 QY 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
 DB 171 TATACGGTGTACGGGATCAAAAGACACTTATTTGGGAGCTCTGGTCTGTGCA 230
 QY 61 GluAlaIleProPheIleLysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 DB 231 GAGGCAATRCCTTTCAATCAAGTTTGGCTTGTGTCCTTGCTATTTATGCTTATGCTT 290
 QY 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
 DB 291 ATTATGCAAGACTAAGTAATATTTTAAGTAAGCAGGCTTATTTATGCAAGTGGAAAG 350
 QY 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
 DB 351 CCCCTTTAAATTTCTTTGCGCTGTGTCGCACTGAATTTATCCCTACGCGATGTTTAA 410
 QY 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProGlyLeuLeuGlyLeu 140
 DB 411 CATCTTAAGAAATTTGCTGACCGCTTTACAGGCCATCTTACCTCCAGGATGCTAGAGATC 470
 QY 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly 160
 DB 471 GTTGCATCTTAAGAAAGTGAACATTTGCTGATTTATGTAATTTGCTGAACTAAGGGGA 530
 QY 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrLysIleHisGlu 180
 DB 531 AGCGTCAGCTATCTCTATATGTTCTGGGATTTGCTAATGAATAACAAAATCCACGAA 590
 QY 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 DB 591 GCAGAGCGTTCTAGCGCTTTTCGGTATCGAGCTAATATTTCTTTACTAGCTTCTGCT 650
 QY 201 ArgAlaIleValAlaSerLysLeuArgAlaSerValSerGlyValAlaAspProTrp 220
 DB 651 CGTGCAATTTGTTGGGCTTCAAAAGTTGAGAGCTTCCGTTTGAAGGGTGAATCCTTGG 710
 QY 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 DB 711 GGAATTTCTTAAAGCTTTTGAATGACTATGACTAATTTGATCGAGCTTGTCTTATGCGC 770
 QY 241 SerTyrTrpTrpIleAsnLysAsnValLeuThrAspProArgPheTyrAsnProGluGlu 260
 DB 771 AGTTCTGGTGAATCAATAGAAAGTATTTGACCGATCTCGCTTCTAATATCAAGAAAGAA 830
 QY 261 MetGlnLysGlyLysGlyValAlaLysProLysMetAsnMetLysAspSerPheLeuTrp 280
 DB 831 ATGCAAAAAGGGAAGAAAGGCTCTAAACCTAATAAGATATGAAGATAGCTTCTCTAT 890
 QY 281 LeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIleAlaTyrGlyIleCys 300
 DB 891 CTGTGATAGATCTCTTATATATTTTAACTCTCTGTTATTTGCTATAGCTATGATTTTGC 950
 QY 301 IleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnTyrProAsnMetAsn 320

DB 951 ATTAACCTAATCGAAGTACTGGAAAAGTCAGCGTGAAGCTGCATATCTAATATGAT 1010
 QY 321 AspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyValValSerValLeuIle 340
 DB 1011 GACTATATGATGTTCAATGGGAACTTCTCTTCTGAGCTGGCGAGATACCGTACTTATC 1070
 QY 341 MetLeuPheValGlyLysValIleAsnValIleAsnLysPheGlyTyrTrpLeuThrGlyAlaLeuVal 360
 DB 1071 ATGCTATTTGTTGGTGAAGTCAATTCGTAAATTTGGATGGTTACTGAGCCCTAGTC 1130
 QY 361 ThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsn 380
 DB 1131 ACTCGTGCATGATTCCTCTTAACAGATACGTTTCTTCTGCTGTTGATCTTAAGAAGC 1190
 QY 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValVal 400
 DB 1191 CAAGCTTCTGGCGTGTGCTATGTTCGTACCACTCTCAAGCTAGCTGTGGTGTTC 1250
 QY 401 GlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeuPheAspSerThrLys 420
 DB 1251 GAGCTATACGAATATTTCTTCAAAATCCAAATATGCTCTCTTTGACTCAACTAA 1310
 QY 421 GluMetAlaTyrIleProLeuAspGlnGluLysValLysGlyLysAlaAlaIleAsp 440
 DB 1311 GAAATGGCTATATCCCTCTTGAACCAAGAGCAAAAGTCAAGGTAAGGCTGATTTGAT 1370
 QY 441 ValValAlaAlaAsnPheGlyLysSerGlyLysAlaLeuIleGlnGlnGlyLeuLeuVal 460
 DB 1371 GTAGTTGGCGCCCGCTTCGAAATCCAGAGAGCTTTAAATCCAAAGGTTTCTGCTT 1430
 QY 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIle 480
 DB 1431 ATCTGGCAATATTTGAGCTATGACCCCTTATCTTGCAGATATCTTCTTTATCATCT 1490
 QY 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
 DB 1491 GCTATTTGTTGTTGTTTTCGCAACTAAGTTAAACAAATCTTATAGCCAGCTGCTCTT 1550
 QY 501 LysGluGlnGluValAlaGlnGluAspSerAlaProAlaSerSer 515
 DB 1551 AAAAGACAAAGAAAGTGGCTCAAGAAAGATTCACTCTCTCTCTCA 1595

RESULT 2
 US-09-938-842A-694
 ; Sequence 694, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Krepes, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRP.300-3
 ; CURRENT APPLICATION NUMBER: US/09/938, 842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227, 866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264, 647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300, 111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 694
 ; LENGTH: 1875
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-694

Alignment Scores:
 Pred. No.: 1.32e-143 Length: 1875
 Score: 1333.00 Matches: 259

Percent Similarity: 69.60%
 Best Local Similarity: 51.80%
 Query Match: 50.68%
 DB: 10
 Gaps: 3

US-09-869-433-2 (1-515) x US-09-938-842A-694 (1-1875)

Conservative: 89
 Mismatches: 140
 Indels: 12
 Gaps: 3

16 LeuTrpProIleHsIStnHISGLuLeuLysValLeuPrometPheLeuMetPhePhe 35
 313 ATTTCGGTGTGAGGTTGCAACCTTGAAAGATATCCCTTTAGAGATGATGTTCTT 372
 36 CysIleThrPheAsnYrThrValLeuAspThrLysAspThrLeuIleValAla 55
 373 TGTATTCCTTTCAATTCACAAATCTGAGGATACAAAGATGCTTGTTGATACGGC 432
 56 ProGlySerGlyAlaGluAlaIleProPheIleLysPheTrpLeuValProCysAla 75
 433 AAGAGAGTTCTCTGAGATTATACCTTCTTGAAGACTTGCGTGAATCTCTATGGCC 492
 76 IleIlePheMetLeuIleYrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPhe 95
 493 ATTGGCTTATGCTCTCTACACTPAACCTTCCAAATGTTCTCTCAAGAGGCTCTGTT 552
 96 TyrAlaValGlyThrProPheLeuIlePhePheAlaLeuPheProThrValIleTyrPro 115
 553 TACACTGTATATGCTCTTTCATCATCTACTTGGGCTTGGTTGCTCATGATACCT 612
 116 LeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProPro 135
 613 CTCGACACTATATTCACCCGGAAGCTCTCGCAATTAAGCTCTTACCAACCTCGGCCA 672
 136 GlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeu 155
 673 AGATTACGGGTCTATTTGCAATATGCGGATTTGGAGTTCTGTTTATGATATG 732
 156 AlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIle 175
 733 GCTGAGCTTTGGGTAAGTGTGTGTCTCACTCTCTTGGGCTTGGCTTCAATCAAGATC 792
 176 ThrLysIleHisGluAlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSer 195
 793 ACAACTGTGATGAGCCCAAGAAATTCATCTTGTGCGACTTGGAGCCCAATGTGCA 852
 196 LeuLeuAlaSerGlyArgAlaIleValTrpAlaSerLysLeuArgAlaSerValSerGlu 215
 853 CTGATTTTCTCAGGAAGAACCGTAATCTCTCTCACTTGAAGAAATCTTGTCCT 912
 216 GlyValAspProTrpGlyLysSerLeuArgLeuMetAlaMetThrIleValSerGly 235
 913 GGAATTACGGCTGGGACAGTTCTTGAAGACCATGATGACATTTGGTGGAAATGGGA 972
 236 LeuValLeuMetAlaSerYrTrpTrpIleAsnLysAsnValLeuThrAspProArgPhe 255
 973 CTCGCCATTTGTCTCTCTATGTTGTGCGTCAATGATATGTTCTCTTCCAAACCGTAGC 1032
 256 TyrAsnProGluGluMetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLys 275
 1033 AAGAAACAGAGAGG-----AAACGGAAGATGGGAACGATG 1068
 276 AspSerPheLeuYrLeuAspArgSerProTyrIleLeuLeuLeuThrLeuLeuValIle 295
 1069 GAAAGCTTGAAGTTCTTGATCATCATCACAATCTTGAAGATCTTGGCTACTTTATGTTG 1128
 296 AlaTyrGlyIleCysIleAsnLeuIleGluValThrTrpLysSerGlnLeuLysLeuGln 315
 1129 GCATACCGTATTAATATCAATCTTGTGAAGTCAATGAAATCAAGACTTAAGCTCAG 1188
 316 TyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPheTrpThrGlyVal 335
 1189 TTCCCTAGCCCGAATGATCTCAGCATTTATGAGAGCTTCTCAACCTGACGGGTGTT 1248
 336 ValSerValLeuIleMetLeuPheValGlyGlyAsnValIleArgLysPheGlyTrpLeu 355

1249 GCAACATTCACAAATGATGCTT---CTCAGCCAAATACGATATTAATGATGTTGGGA 1305
 356 ThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleValPhePheAlaLeu 375
 1306 GTAGCTGCAAGATACACCCCAACTGCTGTGCTATGATGCTGTGCTTCTCTCTA 1365
 376 ValIlePheArgAsnGlnIleAsnSerGlyLeuValAlaMetPheGlyThrTrpProLeuMet 395
 1366 ATATGTTTGGGGCCCATTCGACCACTTGTGCCAAGCTGTGATAGACACCGCTACTT 1425
 396 LeuAlaValValGlyAlaIleGlnAsnIleLeuSerLysSerThrLysTyrAlaLeu 415
 1426 GCAGCTGTATGTCGGTGGCTTCAAGATATCTTCAAGAAAGTGCACCAAGTACAGCTTG 1485
 416 PheAspSerThrLysGluMetAlaTyrIleProLeuAspGlnGluLysValLysGly 435
 1486 TTGACCTTGCACAAAGAAAGGCTTATATCCATGTGATAGACACCAAGTTAAAGGC 1545
 436 LysAlaAlaIleAspValAlaAlaAlaArgPheGlyLysSerGlyAlaLeuIleGln 455
 1546 AAGCTTGCATTAAGTGTGCTGCAACCACTTGGGAAATCGGGAGCTTTATATACG 1605
 456 GlnGlyLeuLeuValIleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIle 475
 1606 CAGTTCATGATCTTATCTTGTGATCATCTAGCAATTCACGCCGTATCTAGAAATATC 1665
 476 LeuLeuPheIleIleAlaIleTrpLeuValSerAlaThrLysLeuAsnLysLeuPheLeu 495
 1666 TTGTTGTTATTTGTCACCTGCGTGTGATGTCGACAGCTAAGTCGCTGGAGGACAGTTC-- 1722
 496 AlaGlnSerAlaLeuLysGlnGluValAlaGlnLysAspSerAlaProAlaSerSer 515
 1723 -----AACGCTTGGCTCTGAAAGAGCTTGAGAGAAATGAGAGAGCTTTCATCG 1776

RESULT 3
 US-09-294-093B-4236
 Sequence 4236, Application US/09294093B
 Patent No. US20010051335A1
 GENERAL INFORMATION:
 APPLICANT: LalGudi, Raghunath, V.
 APPLICANT: Ico, Laura, Y.
 APPLICANT: Sherman, Bradley, K.
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
 FILE REFERENCE: PL-0009 US
 CURRENT APPLICATION NUMBER: US/09/294,093B
 PRIOR FILING DATE: 1999-04-16
 PRIOR APPLICATION NUMBER: 60/082,567
 NUMBER OF SEQ ID NOS: 6207
 SOFTWARE: PERL Program
 SEQ ID NO 4236
 LENGTH: 284
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: misc_feature
 OTHER INFORMATION: Incycle ID No. US20010051335A1 700353546H1
 LOCATION: 132, 274
 OTHER INFORMATION: a, t, c, g, or other
 US-09-294-093B-4236

Alignment Scores:
 Pred. No.: 2,81e-20 Length: 284
 Score: 256.00 Matches: 56
 Percent Similarity: 72.34% Conservative: 12
 Best Local Similarity: 59.57% Mismatches: 26
 Query Match: 9.73% Indels: 1
 DB: Gaps: 0

US-09-869-433-2 (1-515) x US-09-294-093B-4236 (1-284)
 377 IlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThrTrpProLeuMetLeu 396

```
Db      3 TTGTTGGTGAGCCATTGACCTCTTATAGACCAAGTTGGGATACCCCTTTGCTTCG 62
      397 AlAvalValValGlyAlaIleGlnAsnIleLeuSerGlyThrIleTyrAlaLeuPhe 416
      Db      63 GCAGCTTATGTTGGAGCATGAGCAATTTTCTACATAAGAGCAAAATATACAGTCTGTTT 122
      Qy      417 AapSerThrIlyGlnMetAlaTyrIleProLeuAapGlnIleGlnIleValIleGlyIly 436
      Db      123 GATCTTCGNAAGATGCGATCTTCTTGTGATGAGATATGAGGTGAAGTAA 182
      Qy      437 AlAalIleAapValValAlaAlaAapGlyIlySerGlyIlyAlaIleGlnIle 456
      Db      183 GCGGCTATGATGTTGTGTGCAACCCCTGGGGAATCTGGAGGTGCTGTATCAGCAG 242
      Qy      457 GlyLeuLeuValIleCysGlySerIleGlyAlaMetThrPro 470
      Db      243 TTCATGATCCCT-GTCATCGGTTCTCTCCGCAANTCGACACCG 283

RESULT 4
US-09-294-093B-5730
; Sequence 5730, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laigudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294, 093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5730
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700382240H1
US-09-294-093B-5730

Alignment Scores:
Pred. No.: 9,34e-19 Length: 246
Score: 242.00 Matches: 45
Percent Similarity: 74.68% Conservative: 14
Best Local Similarity: 56.96% Mismatches: 20
Query Match: 9.20% Indels: 0
Gaps: 0

US-09-869-433-2 (1-515) x US-09-294-093B-5730 (1-246)
Qy      348 ValIleArgIlyPheGlyTyrIleuThrGlyAlaLeuValThrProValMetValLeu 367
      Db      6 ATCCTCAGAAAGTTGGGTGGGAGCTTGCACATCAACAGCCCTGCAGTCTTACTCTC 65
      Qy      368 ThrGlyIleValPhePheAlaLeuValIlePheArgAsnIleAlaSerGlyLeuValAla 387
      Db      66 ACAAGAGTGGGCTTCTCTCACTGATTTGTTGGTGAACCATGATCCTCTTATGACC 125
      Qy      388 MetPheGlyThrThrProLeuMetLeuAlaValValIleGlyAlaIleGlnAsnIleLeu 407
      Db      126 AAGTTTGGATGACACCTTCTTCCGCGACGCTATGTTGAGCAATGCAAGATTTTC 185
      Qy      408 SerIlySerThrIlyTyrAlaLeuPheAapSerThrIlyGlnMetAlaTyrIlePro 426
      Db      186 AGTAAAGTGCAGAAATACAGTCTGTTGATCCTTGCAAGAGATGGCATATCTCT 242

RESULT 5
US-09-294-093B-1312
; Sequence 1312, Application US/09294093B
```

```
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laigudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294, 093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1312
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700344107H1
US-09-294-093B-1312

Alignment Scores:
Pred. No.: 6,36e-17 Length: 309
Score: 227.50 Matches: 46
Percent Similarity: 66.67% Conservative: 24
Best Local Similarity: 43.81% Mismatches: 32
Query Match: 8.65% Indels: 3
Gaps: 2

US-09-869-433-2 (1-515) x US-09-294-093B-1312 (1-309)
Qy      183 ArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuAlaSerGlyArgAla 202
      Db      1 AAATTTACCCGTTATTTGGCTTGGGCTTAATTTCTCTTATCTTTTCTGGGCTACT 60
      Qy      203 IleValTyrPalaSerIlyLeuArgAlaSerValSerGlyIlyValAapProTyrGlyIle 222
      Db      61 GTGAAGATTTCTCAATTTTCCGCAAGACATTTGGCTCTCGAATGATGCTGGGAAGTA 120
      Qy      223 SerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValIleMetAlaSerTyr 242
      Db      121 TCTTTGAAGAAATGATGACATGATGCTCTCTTGGACTTGTCAATCACTTCATCTAT 180
      Qy      243 TrpTyrIleLeuIlyAsnValIleuThrAapProArgPheTyrAsnProGlnIleMetGln 262
      Db      181 TGGGAGATGAAACAAGTTGTTGATGATGATCTTCACTT-----CCAAAGGCTGATGCT 234
      Qy      263 LyeGlyIlyIlyGlyAlaIlyProIlyMetAsnMetIlyAapSerPheLeuTyrIleuAap 282
      Db      235 AAG---AAGAAAAAGAAAAGCCCTAAACTTGGCATGAAGAGACTCGAAGCTTCTGCTC 291
      Qy      283 ArgSerProTyrIle 287
      Db      292 TCCTCGAGGTATGNG 306

RESULT 6
US-09-738-626-117
; Sequence 117, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
```


Dd	112842	GGCGTGAATTAGGAACCTTTTGGTTCGGCGGAAAAAGATCGTTTGGCAATATGACATCGGT	112901
Qy	100	-----ThrProheuleuilephehealeuilepheProthrValileTyPro	115
Dd	112902	GTGATTTGGCGTACCGCGCGATCATGAGGAGCGGTGCCTTC	112958
Qy	116	LeuAtgAspValLeu-----HisProThGluPheAlaAspArgLeu	129
Dd	112959	TTGGCGCGGATGCTGTTTAAAGCTCTGTCGGCACCCCTGTG	113000
Qy	130	GlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPhe	149
Dd	113001	-----GCTGGGCTATTGGTTCGTTTGGACGCTGC	113030
Qy	150	AlaAlaPheTyValleuAlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrp	169
Dd	113031	GGTGCAGAGATATTCACACAGCATGTGTG-----CCACACAGCTTGATGCATCTTTTGGG	113084
Qy	170	GlyPheAlaAsnGluIle-----ThrylIleHisGlu	180
Dd	113085	GGAAATTACCAACGCGGTATGAGACACTTCAGGCATTCGACACATGAATGCAATCCG	113144
Qy	181	AlaIlyAsArgPheTyAlaLeuPheGlyIleGlyAlaAsnIleSerIleu-----LeuAlaSer	199
Dd	113145	GTTTCTTAACATTTAC-----TTCAATATTTGCATCTCGCATTTGTGTGGCTCTTTATGT	113198
Qy	200	GlyArgAlaIle-----ValTrpAlaSerIlyLeuArgAlaSer	212
Dd	113199	GGTTTCCCATTTGATTAAGGTGTGTGAACCTCGGAGTGTGGCGTCAGAAATTCGCTACGAG	113258
Qy	213	ValSerGluGlyValAspProTrpGly-----IleSerLeuArgLeu	226
Dd	113259	TATGCAGAAAGCATTTGAACCCACGCGACAGATGATGAGAAATCTCTCAACCCTA	113318
Qy	227	-----LeuMetAlaMetThr	231
Dd	113319	ACCGCACAGAAAACCGCGCGCTGACAAATTCATGATGAGACCACTTGCGACGGCCATC	113378
Qy	232	IleValSerGlyLeuValLeuMetAlaSerTyTrpTrpIleAsnIlyAsnValLeuThr	251
Dd	113379	ATCGTGCGTGTGGTGTGCTGATTCGGGATGCCATGAGAAATGAGAGAT	113429
Qy	252	AspProArgPheTyArgAsnProGluIleMetGlnIlyLysIlyGlyAlaIlyAspPolyS	271
Dd	113430	-----GGTGGATTCCTTGCTTACC	113447
Qy	272	MetAsnMetIlyAspSerPheLeuTyTrpLeuAspArgSerProTyIleLeuLeuLeuThr	291
Dd	113448	TGCGCATCTGCTGAGCTCTGTGGTGTATTT-----GTAATTTTGTTTTTCATG	113495
Qy	292	LeuLeuValIleAlaIlyTyGlyIleCysIleAsnLeuIleGluValThrTrpIlySerGln	311
Dd	113496	GTGATGGGCGCTGAGCTACGGCATGTGTGGGACAGATCAAG-----	113537
Qy	312	LeuIlyLeuGlnTyTrpAsnMetAsnAspTySerGluPheMetGlyIlyAsnSerPhe	331
Dd	113538	-----AACATGATCATGTCGGAACATGATGCGGACAAATC-----	113576
Qy	332	TrpThrGlyValIleValSerValLeuIleMetLeuPheValIleGlyIlyAsnValIleArgIlyS	351
Dd	113577	---AAGACATGATTTGGTCTTGTTGGCTTCATTTTGGGACAGATTTTGGCGCTG	113633
Qy	352	PheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal	371
Dd	113634	TTTAACTG-----ACGGGCAATCGGT	113654
Qy	372	PhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeu-----ValAlaMetPhe	389
Dd	113655	ACCTGACACTGCTGTT-----CAGGTCCTGCGGAGATTGGAAGCATCGGCTTACC	113705
Qy	390	GlyIlyThrTrpProLeuMetLeuAlaValAlaValGlyAlaIleGlnAsn-----IleLeu	407
Dd	113706	GGATTTCCCTCGATCATGATTTATATTTTGGCGTATGTAACCTGCTGATTAAT	113765

QY 408 SerlySerThrLysTyrAlaLeuPheAspSerThrLysGluMetAlaTyrIleProLeu 427
 DB 113766 TCGGCTCTGCGATGCGAGCTG-----ATGGCTGGCGTTCGTC 113807
 QY 428 AspGlnGluInhLysValLysGlyLysAlaAlaIleAspValAlaAlaArgPheGly 447
 DB 113808 CCGATGTTGCGTCTGGGTATGAACATCATTCAAGAGCATTCGGGTGGGT 113867
 QY 448 LysSerGlyGlyAlaLeuIle-----GlnGlnGlyLeuValIleCysGlySer 464
 DB 113868 GACTCGGCACTCAGGATGATCACCGCTGATCCGATGATGATGATCCTCGGTGG 113927
 QY 465 IleGlyAlaMetThrPro-----TyrLeuAlaValIleLeuLeuPhe 478
 DB 113928 CTCGCTGATACGACCGGATGACGATTAGGACCTTGATGATCAAGCTTATCCATT 113987
 QY 479 IleIleAlaIleTyrLeuValSerAlaThrLysLeuAsnLysLeuPheAlaGlnSer 498
 DB 113988 GTGATCCCTTCTTGCGTACGCTGGGCTACATGTTGGCAATTGGTTCTACGCCGATTG 114047
 QY 499 AlaLeu 500
 DB 114048 CCGCTT 114053
 RESULT 8
 US-09-815-242-6064
 ; Sequence 6064, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 6064
 ; LENGTH: 1389
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1389)
 ; US-09-815-242-6064
 Alignment Scores:
 Pred. No.: 1.34e-05 Length: 1389
 Score: 138.50 Matches: 111
 Percent Similarity: 37.55% Conservative: 70
 Best Local Similarity: 23.03% Mismatches: 172

Query Match: 5.27% Indels: 130
 DB: 9 Gaps: 24
 US-09-869-433-2 (1-515) x US-09-815-242-6064 (1-1389)
 QY 69 TrpLeuValValProCysAlaIle-----IlePheMetLeuIleTyrAlaLys 84
 DB 109 TGGATATCACTAGCTTGTCTTCTTCTTCCAAAGCCAAACAGGAGCGCTTT---GGATTAA 503
 QY 85 LeuSerAsnIle-LeuSerLysGlnAlaLeuPheTyrAlaValGlyThr----- 100
 DB 169 ACCGTAATCTCAATAAATCGGTTTAATTATTAACGATCAACCTTTTATTAACC 228
 QY 101 -----ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeu 117
 DB 229 GCATTACCTTCCTTCTTGGCGATATGCGGTTTCTTATGCTTATGATGCTTAT 288
 QY 117 GAAPValLeuIleProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLe 137
 DB 289 TTGGTGAGACCGCGATGAGCGTTT---AGTACTGCAATCTGATTATTCCTTGGCTG 347
 QY 137 uLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaG 157
 DB 348 GCTCGGATGCGCGTGCAGAAATCCGAATACTCTTTGGGATATTTATCGTTATCGCTT 407
 QY 157 uLeuTrpGlySerValMetLeuSerLeuMetPheTrpGly-----PheAlaAsnGlu 175
 DB 408 GCTATGCGGT-----TTTGAGGTGCAGAACTTCTTCGACCAT 446
 QY 175 ethrLysIleIle-----GluAlaLysArgPheTyrAlaLeuPheGlyIleG 191
 DB 447 GGGCATATGATGTTCTTCTTCCAAAGCCAAACAGGAGCGCTTT---GGATTAA 503
 QY 191 yAlaAsnIleSerLeuLeuAlaSerGlyArgAlaIleValITrpAlaSerLysLeuArgAl 211
 DB 504 TGGCGGATTA-----GGAACATTAGGTGT 527
 QY 211 aSerValSerGluGlyValaAspProTrpGlyIleSerLeuArgLeuLeuMetAlaMet 231
 DB 528 AAGGTATGACAGCTGGTGCACCGCTGCTGATTTTGTATGCTGATTTGCTTCGG 587
 QY 231 rIle-----ValSerGlyLeuValLeuMetAlaSerTy 242
 DB 588 CGTCAATGCGTACCGGACGCGACCGGTTGGTGATGATCGCTGCGAATCCGATGATGAT 647
 QY 242 rTrp-----TrpIleAsnLysAsnValLeuThr 251
 DB 648 TTGGGTACCGCTACGCGATGTCACAGATCCGCGATGTCAGGATGATGATATC-- 705
 QY 251 rAspProArgPheTyrAsnProGluMetGlnLysGlyLysValAlaLysProLy 271
 DB 706 -----GCCAGTTCACG 716
 QY 271 smetAanMetLysAspSerPheLeuTyrIleAspArgSerProTyrIleLeuLeuThr 291
 DB 717 CGCTCAATGCGGACGACGCTCCCTGTTACACGCTG---CATCTGCGCTGCTGAG 773
 QY 291 rLeuLeuValIleAla---TyrGlyIleCysIleAsnLeuIleGluValThrTrpLysSe 310
 DB 774 CTGCTTACCTTGCACGCTTCGTTGCTTATCGGT---TTTTCGCGGCTTTTGCAT 830
 QY 310 rGlnLeuLysLeuGlnTyrProAsnMetAsn-----AspTyrSerGluIup 325
 DB 831 GCTGGCAAAAACCAAGTCCCGGATGGAATATTCGCGCTGCGCTTCTTGGCCCAT 890
 QY 325 emetGlyAsnPheSerPheThrGlyValValSerValLeuIleMetLeuPheValG 345
 DB 891 TATCGGT-----GCCATCGCGGCTGCTG 917
 QY 345 yGlyAsnValIleArgLysPheGlyTyrLeuThrGlyAlaLeuValThrProValMetVa 365
 DB 918 TGGTGTATTTCCGATGATGATTCGGCGGCGGCGGATGATGATCAACTTATTTTAT 977

```

QY      365  lLeuLeu-----ThrIly-----11 370
      : :::::
Db      978  GGGCAATTTCAGTCCCTGCTGTCTTCACTTACCGGGACAGAGCTCCGGTAATTTCAAT 1033
QY      370  evalaPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheG1 390
      : :::::
      : :::::
Db      1038  CGCCTTTTACGGCGATATTATGGGCGTGTCTGACTGGGGGTCTGGGAAT-----GG 1091
QY      390  yThrThrProLeuMetLeuAlaValAlaGlyAlaIleGlnAlaMetIleLeuSerLyse 410
      : :::::
      : :::::
Db      1092  TTCTACCTTCCAGATGATCGCGGTATCTTTGCCAGAA-----1133
QY      410  rThrLyThr-----AlaLeuPheAspSerThrLySGlnMetAlaTyrIleProLeuAs 428
      : :::::
      : :::::
Db      1132  -ACCATTTATTCGGGTAAAGATCAAAAGCGCGTGTATGATGACAGACTCAT-----1179
QY      428  pGlnGlnGlnLysValLysGlyLysAlaAlaIleAspValValAlaAlaArgPheGlyLy 448
      : :::::
      : :::::
Db      1180  -AAAGAACCCGTCACCGGAACGCGCGCGCTCTGCTTATCTCAGCC--ATTGGCGCG 1235
QY      448  sSerGlyGlyAlaLeuIleGlnGlnGlyLeu-----IleValIleCysGlySer-- 464
      : :::::
      : :::::
Db      1236  AGTGGGGCGCTTTTATTTCGCGACGGCGTTTGGCATATGCTCAATATGACCGGCTCTCC 1295
QY      465  -IleGlyAlaMetThrProTyrLeuAlaValIleLeuPhePheIleIleAlaIleTyrLe 484
      : :::::
      : :::::
Db      1296  GGTGCGGCGCATGAAGAAGTGTTTTATCTTTCATCATCGTTGTGTGCTGACACTGGCT 1355
QY      484  uVal 485
      : :::::
      : :::::
Db      1356  GGT 1359

```

```

RESULT 9
US-09-770-961-833
Sequence 833, Application US/09770961
Publication No. US2003011563A1
GENERAL INFORMATION:
APPLICANT: Goriach, Jörn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameeka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kricker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hubban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: thaliana
FILE REFERENCE: 2026 (PRA-015PRV)
CURRENT APPLICATION NUMBER: US/09/770,961
CURRENT FILING DATE: 2001-01-26
PRIORITY APPLICATION NUMBER: 60/178,466
PRIORITY FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 833
LENGTH: 484
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-961-833
Alignment Scores:
Pred. No.: 2.92e-06
Length: 484

```

Score:	138.00	Matches:	25
Percent Similarity:	76.74%	Conservative:	8
Best Local Similarity:	58.14%	Mismatches:	10
Query Match:	5.25%	Indels:	0
DB:	11	Gaps:	0

```

OY      16 LeuTTPProIIeHishrHxhSglueLysValLeuPromeCtPheLeuMeCPhePhe 35
          ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      355 ATTTTCGGTGGTGAGGCTTAGCACTGTGAAGAAGATGTCTTTTAGGGCTAATGTCTTT 414
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      36 CysIleThrPheaSnTyThrValLeuArgAspThryLysApPrhLeuIleValGIyAla 55
          ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      415 TGCATCCCTTTCAATTACACAATCTTAGGAGCACGAAGATGTTTTGTGTGACGGCT 474
          |||||
OY      56 ProGIlySer 58
          |||||
Db      475 AAGGAAGT 483

RESULT 10
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIOR FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature

```

```
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10091)..(10091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
```

```

Oy 270 -----ProlysmetAamMetLysAspSer-PheLeuTyr----- 280
Db 1778731 ATATTAAAGTAAATTAACCTTAACCTTAAGCATTTTGGTTACTCTCTTGTGTTTAATTCA 1778672
Oy 281 ----LeuAspArgSerProTyrIleLeuLeuLeuThrLeuValIleAlaTyrGlyI 299
Db 1778671 GGCCCTACGTTATACCTTTTATCTGCATTTAAGCACAATATGGTAATT -----AT 1778621
Oy 299 eCySIIeasLeuIleGluValThrTrpLysSerGlnLeuLysLeuGlnIlyrProAsnMe 319
Db 1778620 TTCAAGC-AATTACTGTCAACTCAAGTTTAAACC-----TATGCTTATGA 1778577
Oy 319 LAsAsPTySerGlnuPheMetGlyAenPheSerPhe----- 331
Db 1778576 ACAAGAACATACACTGCTGTTTAGCGGATGAGACTGTGCTTATTGGGCTGTGTGTTC 1778517
Oy 332 -TrpThrGlyValValSerValLeuIleMetLeuPheValGlyGlyAsnValIleAlaGly 351
Db 1778516 TTGGGCTCCGCTTGTGGGTTTATTATTAATGCGGATTTCTAAAGGGGGAACATACCCGTGA 1778457
Oy 351 sPheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuThrGlyIleVal 371
Db 1778456 ATTT-----ATTTTGGCGATTAAGTATTC-----AGTTTATTTGGTATTTT 1778412
Oy 371 lPhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyTh 391
Db 1778411 ATGGTT-----ACTGTTTTGGTAA 1778391
Oy 391 rThProLeuMetLeu-----AlaValValGlyAlaIleGlnAsnIleLeuSerLy 409
Db 1778390 TACAGCAGTATGCGTAAATGATGCGATGTCGCGGGCGGCTTGGCGAATTAATTTCT- 1778333
Oy 409 sSerThrLysTyrAlaLeuPheAspSerThrLysGlnMetAlaTyrIleProLeu----- 427
Db 1778332 -TCCCCGAAATTTTATTAATTT-----AAATTTTAATTAATTCGCTTACCAAC 1778283
Oy 427 ----- 427
Db 1778282 AATAACAGCGCTTGTGAGCTTATAGTATTAATGTTTATTTATCACTTCAGCCGATTC 1778223
Oy 428 -----ArgGlnGlu----- 430
Db 1778222 AGTATTATTAATGTGTTAAATAACATTCGATCGTGTGAATAAAAGTTTACCTGCGCTGCGTG 1778163
Oy 431 -GlnLysValIleGlyLysAlaAlaIleAspValValAlaAlaArgPheGlyLysSerGly 450
Db 1778162 GCAGGCCATTAATGCGGGAACCTTAATATGCCGTTGTTCGATTCGTTATGCAATCTGG 1778103
Oy 450 yGly-----AlaLeuIleGlnGlnGlyLeuLeuValIleCyGlySerIleGlyAlaMetTh 469
Db 1778102 TGACCTTGCTAATTTGCAACAACATGACATTAATGTGGCC-----TT 1778061
Oy 469 rProTyrLeuAlaValIleLeuLeuPheIleIleAlaIleTrp-----IleValSer-- 486
Db 1778060 GCCCTTTGCTTAATGATGTTGGTAATGCTTTTAATGGAAGGCTTAATTCGGA 1778001
Oy 487 -----AlaThrLysLeuAsn 491
Db 1778000 TAAAAAATATTTTCTACTAAAGCAAT 1777973

RESULT 11
US-09-815-242-6047
; Sequence 6047, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```


CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 9794
 LENGTH: 1188
 TYPE: DNA
 ORGANISM: Salmonella typhi
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1188)
 US-09-815-242-9794

Alignment Scores:
 Pred. No.: 0.00127
 Score: 120.50
 Percent Similarity: 34.55%
 Best Local Similarity: 19.92%
 Query Match: 4.58%
 DB: 9
 Gaps: 20

US-09-869-433-2 (1-515) x US-09-815-242-9794 (1-1188)

28 LeuProMetPheLeuMetPhePheCysAlIlePhePheAsnTyrrHrValLeuArgAspThr 47
 7 TTAATGATGCTCTCGCTTATTTGCACTGACAAATAGCTTCCTGCTATT--GGCAGC 63
 48 LysAspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIleProPheIleLys 67
 64 ACTGAGTTCGTAATGTCGCGC-----CTGGTCCGACCATTTGCA 102
 68 PheTrpLeuValVal-----ProCysAlaIleIlePheMetLeuIleTyrrAlaLysLeu 85
 103 CAACAACCTGCTGCTGCTTACCTTCTGCGGAATGCTGTTTAACTAT----- 153
 86 SerAsnIleLeuSerLysGlnAlaLeuPheTyrrAlaValGlyThrProPheLeu----- 103
 154 -----GCACTGGGTGGCTATTGGCGCGCGCTGCTGCTAACCGCT 192
 104 -----IlePhePheAlaLeuPheProThrValIleTyr 114
 193 CTGACCGGAGCGCTCGCGCGTAACAGTTACTGCTGCGCTGAAG-----GTCCTGTTT 246
 115 ProLeuArgAspValLeuHisProThrGluPheAlaAspArgLeuGlnAlaIleLeuPro 134
 247 ACG-----GCGGGGAATTGCTGCGCGTGCAGAGCT 276
 135 ProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThrPheAlaIlePheTyrrVal 154
 277 CCGGCTATATGACATTTAATCGCC--GCCGCTGCTGACGAGGCGCTGCGCAGCGCGCTG 333
 155 LeuAlaGluLeuTrpGlySerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGlu 174
 334 TTCTTCGTGATGCTCTACCATCGCAAGCTG----- 369
 175 IleThrLysIleHisGlyAlaLysArgPheTyrrAlaLeuPheGlyIleGlyAlaAsnIle 194
 370 GTGCCGAAAGAAAGAGCGCTTCTGCTATCGCTATTTATGTTGCT--GGTTTAAAGATA 426

195 SerLeuLeuAlaSer-----GlyArgAlaIleValTrpAlaSer 207
 427 GCGCTGTACAGAGTGTCCGCTGGGACATTTATCGTACAGATTTGGCTGCGC--- 483
 208 LysLeuArgAlaSerValSerGlyValAlaAspProTrpGlyIleSerLeuArgLeuLeu 227
 484 -----GAGACCTTC 492
 228 MetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyrrTrpIleAsnLys 247
 493 CTGGCGTATCATTGATGCGGTGATGCGCTGATGCGAGTCACTGATGCTGCGCC 552
 248 AsnValLeuThrAspProArgPheTyrrAsnProGluGluMetGlnLysGlyLysGly 267
 553 AATATT-----CTGCTCGGCGGCC 573
 268 AlaLysProLysMetAsnMetLysAspSerPheLeuTyrrLeuAspArgSerProTyrrIle 287
 574 GCC-----ACATTCGTGACCAAGTGAAGTACTGACTCATCTCGTCTGCTG 621
 288 LeuLeuLeuThrLeuLeuValIleAlaTyrrGlyIleCysIleAsnLeuIleGluValThr 307
 622 CTGATTACCGCCGTTACGGCGCTGCGCTACGGCGGATTC----- 663
 308 TrpLysSerGlnLeuLysLeuGlnTyrrProAsnMetAsnAspTyrrSerGluPheMetGly 327
 664 -----ACGCGATTACCTTCCTGCGCGCGCATGATGAGATTCGCGGATTTCTCCG 717
 328 AsnPheSerPheTrpThrGlyVal-----ValSerValLeuIleMetLeuPheVal 344
 718 GCAAGATAAGCTGAGTCTACTGAGGTGTTATGGGTATCAGTGGTATCGCAATATCTGG 777
 345 GlyGlyAsnValIleArgLysPheGlyTyrrLeuThrGlyAlaLeuValThrProValMet 364
 778 GCGCGTAACCTGCGGATGAGCATGCTG-----GCCGTGCCGCGCTCTG 819
 365 ValLeuLeuThrGlyIleValPhePheAlaLeuValIlePheArgAsnGlnAlaSerGly 384
 820 AATTCATTTCGCGCGCGCTGTTTCTCTGCTGATGCTTCCAGAGTACCGCTCCAGC 879
 385 LeuValAlaMetPheGlyTyrrThrProLeuMetLeuAlaValValGlyAlaIleGln 404
 880 CAGTATCCCGCGCTGCTACCATTTGCTGATGGGATCTTTGCTCGTAACGTACCG 939
 405 AsnIleLeuSerLysSerThrLysTyrrAlaLeuPheAspSerThrLysGluMetAlaTyrr 424
 940 GGGTTG-----CAGGTCTAC 954
 425 IleProLeuAspGlnGluGln-----LysValLysGlyLysAlaAlaIleAsp 440
 955 GTGGTACAGAAAGCGAACAATTACAGCTTAATGCCGTGATGCGCTGACGTCTAAC 1014
 441 ValValAlaIleArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuVal 460
 1015 ATTCCGCGCTTAAATGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074
 461 IleCysGlySerIleGlyAlaMetThrProTyrrLeuAlaValIleLeuLeuPheIleIle 480
 1075 CATTAACGTTTG-----GCCGAGACGCGCTGATGCGCCCTCATTTCTGCTGCTGCT 1128
 481 AlaIleTrpLeuValSerAlaThrLysLeuAsnLys 492
 1129 TTCTCTGATGAGCGCTGACGAGACGCTTGTATTA 1164

RESULT 14
 US-09-815-242-8073
 Sequence 8073, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haeselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 8073
 LENGTH: 1278
 TYPE: DNA
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1278)
 US-09-815-242-8073

Alignment Scores:
 Pred. No.: 0.00142 Length: 1278
 Score: 120.50 Matches: 94
 Percent Similarity: 37.27% Conservative: 86
 Best Local Similarity: 19.46% Mismatches: 160
 Query Match: 4.58% Indels: 143
 DB: 9 Gaps: 19

US-09-869-433-2 (1-515) x US-09-815-242-8073 (1-1278)
 QY 9 PheGlyLysLeuArgSerPheLeuThrProIleHisGluLeuLysValLeu 28
 DB 19 TTTGGAGCGGAAATTATATATCCGCCCATGCTTGCCATACACGCGGTCAAAATATG 78
 QY 29 PrometPheLeuMetPhePheCysIleThrPheLeuThrValLeuArgAspThrLys 48
 DB 79 TGGATTGGTATGCTAGGCTTGCCCTTACA-----GGCATATTACTCCCTTTATT 129
 QY 49 AspThrLeuIleValGlyAlaProGlySerGlyAlaGluAlaIle----- 63
 DB 130 ACTGTTATTGTTGTCATTTTATGATGAGGTGTAAGGTGTAAGCAATCGTATACAT 189
 QY 64 ProPheIleLysPheThrLeuValAlaProCysAlaIleIlePheMetLeuIleThrAla 83
 DB 190 CCATGGTTCGGGTTATTTTGGCTGTC-----GTGATTATCATGCTCATCGGACA 240
 QY 84 -----LysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaVal 98
 DB 241 TTTTACGCTATTCACGCTGCGCAATGTCGG-----TACGAAT 282
 QY 99 GlyThr-----ProPheLeuIlePhePheAlaLeu 108
 DB 283 GGTACAAACACATTTTACTGTCGACAAACATGACCTTAATATATTCGACGACAGTC 342
 QY 109 PheProthrValIleThrProLeuArgAspValLeuHisProthrGluPheAlaAspArg 128
 DB 343 TTTTGGCATGCTTTACTGATTAAGT-----TTAAATCATGCAAAATGCTGATATAT 396
 QY 129 LeuGlnAlaIleLeuProGlyLeuLeuGlyLeuValAlaIleLeuArgAsnTrpThr 148

DB 397 TTAGTAATATATTAACCGCTTACTATTAATGTCGCTATTA-----AGT 447
 QY 149 PheAlaAlaPheThr----- 153
 DB 448 ATTGCTGTCAATTTCAACCGTGAATCTGACCTAGTGCACCTAAGATTAATATATACA 507
 QY 154 -----ValLeuAlaGluLeu 158
 DB 508 CATCCTTCATTTGAGGAAGTTTGAAGGCTATTTTACATGACCTTGTTGCTGCGTTA 567
 QY 159 TrpLysSerValMetLeu----- 164
 DB 568 GCTTTTCCGAGTACATGTCATGCTATTAAGTTAAAGGCTCACAGATCGCAGAAA 627
 QY 165 SerLeuMetPheThrPheGlyPheAlaSerGluIleThrLysIleHisGluAlaLysArgPhe 184
 DB 628 ATATTAAAGTATGCTCTCTTTTCAGATTATTCACCTATTTACTGTGTATGATTATAC 687
 QY 185 TyrAlaLeuPheGlyIleGlyAlaAsnIleSer-----LeuLeuAlaSerGlyArgAla 202
 DB 688 TTTCGTTGGCATATGATGATGCTGCTCACAGCTTCAGAAACTTAAAGATGTTACTGAT 747
 QY 203 IleValTrpAlaSerLysLeuArgAlaSerValSerGluGlyValAlaAspProThrPheGly 222
 DB 748 ATATTGACATACATATTCATACCGGTATTTGTTGCTTCGTTAACCTCGTATTGGAATG 807
 QY 223 SerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAlaSerTyr 242
 DB 808 ACGGTT---ATTCTTGATGCTCAACCAATGATATAGACTCGTC---AAAGCTTGTCGA 861
 QY 243 TrpThrIleAsnLysAsnValIleThrAspProArgPheTyrAsnProGluGluMetGln 262
 DB 862 ACATTTCATAGAACACGTC----- 882
 QY 263 LysGlyLysLysGlyAlaLysProLysMetAsnMetLys----- 275
 DB 883 -----CCTAAGTTTCTTAATAAAATATTCGACCTGTTTCTCT 921
 QY 276 ---AspSerPheLeuThrLeuAspArgSerProThrIleLeuLeu-----LeuThr 291
 DB 922 ATCATAGATTTTATTTATACACACTTGTTAGAAATGATTTTAAATGCTGTACCA 981
 QY 292 LeuLeuValIleAlaTyrGlyIleCysIleAsnLeuIleGluValThrTrpLysSerGln 311
 DB 982 TTGTTGACATTAATATATCCCGTCTCATTTGCACTTACTGATATCATTT----- 1032
 QY 312 LeuLysLeuGlnTyrProAsnMetAsnAspTyrSerGluPheMetGlyAsnPheSerPhe 331
 DB 1033 -----GCTAACATGTTTAGACATTCAGATTCAAGTTGGGCTAT 1071
 QY 332 TrpThrGlyValAlaSerValIleLeuMetLeuPheValGlyLysValIleArgLys 351
 DB 1072 CGACTCGCAACGTGTTATTTACATGATTAATTTCAATTTTA-----CAAAATCTAAATAGT 1125
 QY 352 PheGlyTrpLeuThrGlyAlaLeuValThrProValMetValLeuLeuThrGlyIleVal 371
 DB 1126 TTCACTTATTAACGCGGTGTTATTTGAATGTTTATGATGTTTACTTACGACGATATC 1185
 QY 372 PhePheAlaLeuValIlePheArgAsnGlnAlaSerGlyLeuValAlaMetPheGlyThr 391
 DB 1186 GATTTACTTGGCTGTA----- 1203
 QY 392 ThrProLeuMetLeuAlaValAlaValGlyAlaIleGlnAsnIleLeuSerLysSerThr 411
 DB 1204 ---CATTCATTTCTTTGGCAATCATGTTTCAATCATCATGATGATGATTATACCGCTCG 1260
 QY 412 LysTyrAla 414
 DB 1261 AAACAGCG 1269
 RESULT 15
 US-10-085-959-109/c


```

: Sequence 109, Application US/10085959
: Publication No. US20030165870A1
: GENERAL INFORMATION:
: APPLICANT: Blattner, Frederick R.
: APPLICANT: Welch, Rodney A.
: APPLICANT: Burland, Valerie D.
: TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
: FILE REFERENCE: 960296.97648
: CURRENT APPLICATION NUMBER: US/10/085,959
: PRIOR FILING DATE: 2002-03-01
: PRIOR APPLICATION NUMBER: 60/242,412
: NUMBER OF SEQ ID NOS: 255
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 109
: LENGTH: 76804
: TYPE: DNA
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (39751)..(39751)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (49372)..(49372)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (66539)..(66539)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73693)..(73693)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73705)..(73705)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73737)..(73737)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73759)..(73759)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73761)..(73761)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73766)..(73766)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73774)..(73774)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73781)..(73781)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73788)..(73788)
: OTHER INFORMATION: Unsure
: NAME/KEY: misc feature
: LOCATION: (73802)..(73802)
: OTHER INFORMATION: Unsure
US-10-085-959-109

Alignment Scores:
Pred. No.: 0.886 Length: 76804
Score: 120.50 Matches: 109
Percent Similarity: 36.65% Conservative: 79
Best local Similarity: 21.25% Mismatches: 173
Query Match: 4.58% Indels: 152
DB: 12 Gaps: 27

US-09-869-433-2 (1-515) x US-10-085-959-109 (1-76804)

09 94 LeuheaYrYrIAVaIGYThrProphelaLeuIleehhehalaLeuPhProthValIle 113
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
71299 ATATTTCGGTGGGACGACGAGTTATTACTTGATTCGTTGATATCT 71240

```

```
OY 403 IlegInAsnIleLeuSerLysSerThrLysTyr-----AlaLeuPheAspSer 418
|||
|||
Db 70297 ATTATCCCCGTTTCTTCTTGACGCGGACTATGATGCTTACTGGCCTTCTAATCTTT 70238
OY 419 ThrLysGluMet-----AlaTyrIlePro----- 426
|||
|||
Db 70237 ACTCAGCAGGTTTGTGTCAGGGGATGTGCTCTGTGTCGCGAAATAATATCAGTGTAT 70178
OY 427 LeuAspGlnGluGlnLysValLysGlyLysAlaAlaIleAspValValAlaAlaArgPhe 446
|||
|||
Db 70177 TTTAATGTGACGACGCGGCGGAGACTGGGGTTTATTATATATGTA----- 70130
OY 447 GlyLysSerGlyGlyAlaLeu-----IlegInGlnGlyLeuLeuValIleCysGlySer 464
|||
|||
Db 70129 GGTTCACCTGTGTGACGACATGCGCCTATTCTTGGTCAGTGTGTGCGTCATACAGT 70070
OY 465 IlegIAlaMetThrProTyrLeuAlaValIleLeuLeuPheIleIleAlaIle----- 482
|||
|||
Db 70069 CTTGGTACGACATGTGTTCTTGTCTTATCTTAACATTTGTTGTTCTTCTTATT 70010
OY 483 -----TyrLeu-ValSerAlaThrLysLeuAsnLys 492
|||
|||
Db 70009 GGTTTGATATGCCATCACTGATACAGAGATGATCCATCCAGAACTGCGCTGAATAT 69950
OY 492 sLeuPheLeuAlaGlnSerAlaLeuLysGluGlnIle 504
|||
|||
Db 69949 GATACTGTCGATGAGAAACCTTTTATGAGCAGAA 69913
```

Search completed: November 25, 2003, 14:53:12
Job time : 1979 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:50:16 ; Search time 28 Seconds
(without alignments)
1768.817 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515
Sequence: 1 MTKTEKPEFKLRSFLWPIH.....AQSALKEQVAVQSDAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: PIR 76: *
2: pir1: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313	60.8	515	2	ADP, ATP carrier p
2	224	43.5	515	2	ADP/ATP translocas
3	33	6.4	528	2	probable adp/atp t
4	33	6.4	529	2	ADP, ATP carrier p
5	13	2.5	501	2	ADP, ATP carrier p
6	13	2.5	501	2	ADP, ATP carrier p
7	11	2.1	589	1	ADP/ATP translocas
8	11	2.1	589	1	ADP/ATP translocas
9	11	2.1	624	2	adenine nucleotide
10	11	2.1	624	2	adenine nucleotide
11	11	2.1	624	2	adenine nucleotide
12	10	1.9	511	2	ADP/ADP-transporte
13	10	1.9	511	2	ADP/ADP-transporte
14	9	1.7	463	2	ADP, ATP carrier p
15	9	1.7	463	2	ADP, ATP carrier p
16	9	1.7	463	2	ADP, ATP carrier p
17	8	1.6	498	1	ADP/ATP translocas
18	8	1.6	498	1	ADP/ATP translocas
19	8	1.6	498	1	ADP/ATP translocas
20	8	1.6	498	1	ADP/ATP translocas
21	8	1.6	498	1	ADP/ATP translocas
22	8	1.6	498	1	ADP/ATP translocas
23	8	1.6	498	1	ADP/ATP translocas
24	8	1.6	498	1	ADP/ATP translocas
25	8	1.6	498	1	ADP/ATP translocas
26	8	1.6	498	1	ADP/ATP translocas
27	8	1.6	498	1	ADP/ATP translocas
28	8	1.6	498	1	ADP/ATP translocas
29	8	1.6	498	1	ADP/ATP translocas

30	8	1.6	472	2	S74886
31	8	1.6	480	2	AD1932
32	8	1.6	485	1	ANHU
33	8	1.6	485	1	TVMVT3
34	8	1.6	1772	2	T36105
35	8	1.6	1772	2	T36105
36	7	1.4	1964	2	T09059
37	7	1.4	80	2	S10464
38	7	1.4	81	1	T5P2A
39	7	1.4	82	2	UC5892
40	7	1.4	83	1	NSCH4
41	7	1.4	108	2	E90163
42	7	1.4	110	1	C64430
43	7	1.4	116	2	G70831
44	7	1.4	146	2	E54403
45	7	1.4	146	2	E54403
46	7	1.4	152	2	H71129
47	7	1.4	166	2	D86026
48	7	1.4	172	2	S36007
49	7	1.4	173	2	T11831
50	7	1.4	176	2	AB3363
51	7	1.4	191	1	JH0585
52	7	1.4	193	2	I50693
53	7	1.4	220	2	AB3321
54	7	1.4	235	2	AB2598
55	7	1.4	239	2	D71856
56	7	1.4	240	2	A64553
57	7	1.4	245	2	G82068
58	7	1.4	252	2	F64934
59	7	1.4	252	2	H90935
60	7	1.4	252	2	D85784
61	7	1.4	259	2	S50337
62	7	1.4	260	2	AE2840
63	7	1.4	263	2	G97617
64	7	1.4	264	2	T47468
65	7	1.4	264	2	F88504
66	7	1.4	268	2	H70118
67	7	1.4	268	2	D97548
68	7	1.4	272	2	C87711
69	7	1.4	275	2	AH1435
70	7	1.4	275	2	A11077
71	7	1.4	276	2	C75414
72	7	1.4	277	2	T29611
73	7	1.4	284	2	C97380
74	7	1.4	295	2	B69653
75	7	1.4	296	2	T34256
76	7	1.4	301	2	T33702
77	7	1.4	306	2	AG2698
78	7	1.4	306	2	G97480
79	7	1.4	309	2	T44663
80	7	1.4	315	2	AB3435
81	7	1.4	317	2	F96656
82	7	1.4	320	2	AD2066
83	7	1.4	320	2	B97206
84	7	1.4	322	2	T06955
85	7	1.4	323	2	H91279
86	7	1.4	323	2	H86120
87	7	1.4	324	2	S56457
88	7	1.4	324	2	T27302
89	7	1.4	329	2	T16469
90	7	1.4	330	2	C69329
91	7	1.4	335	2	H96752
92	7	1.4	336	2	AB1324
93	7	1.4	336	2	AD1695
94	7	1.4	336	2	H82040
95	7	1.4	346	2	T11325
96	7	1.4	348	2	F83993
97	7	1.4	352	2	AG0762
98	7	1.4	354	2	S70595
99	7	1.4	364	2	F95177
100	7	1.4	364	2	H98043

phytoene dehydroge
hypotheical prote
angiotensin precu
transforming prote
probable large gly
notch4 - mouse
fasciculin-like p
short toxin 2 prec
cobrotoxin b - Ch1
short neurotoxin 1
nonhistone chromos
1S1 ribosomal prot
ribosomal protein
hypotheical prote
hemoglobin C beta
hypotheical prote
hypotheical prote
NADH2 dehydrogenas
NADH2 dehydrogenas
peptidylprolyl iso
ribosomal protein
interferon - chick
conserved hypothe
hypotheical prote
probable holocyto
probable holocyto
ribosomal large su
hypotheical prote
hypotheical prote
hypotheical prote
cytochrome-c oxida
ABC transporter, m
probable ABC trans
hypotheical prote
protein B0361.2 [i
glucosamine-6-phos
lipase-protein 11
hydrolyase, alpha/b
PTS system, mannos
PTS system, mannos
hypotheical prote
hypotheical prote
probable ATP-bind1
transmembrane lipi
hypotheical prote
hypotheical prote
cation efflux syst
hypotheical prote
hypotheical prote
prochome ix farne
cobalamn biosynth
methyl-accepting c
probable cytochrom
probable transport
probable transport
hypotheical 34.0K
hypotheical prote
hypotheical prote
immunogenic protei
hypotheical prote
transcription regu
genral secretion
NADH2 dehydrogenas
C4-dicarboxylate t
probable membrane
NADH2 dehydrogenas
transulfuration en
O-succinylhomoseri

ALIGNMENTS

RESULT 1

ADP, ATP carrier protein CP0408 [imported] - Chlamydia pneumoniae (strain CWL029 at E72089
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C/Accession: E72089, D81580
 R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A/Reference number: A72000; MUID:99206606; PMID:10192338
 A/Accession: E72089
 A/Molecule type: DNA
 A/Residues: 1-515 <ARN>
 A/Cross-references: GB:AE001619; GB:AE001363; NID:g4376620; PIDN:AD18495.1; PID:g437662
 A/Experimental source: strain CWL029
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: D81580
 A/Molecule type: DNA
 A/Residues: 1-515 <REA>
 A/Cross-references: GB:AE002202; GB:AE002161; NID:g7189324; PIDN:AA38252.1; PID:g718933
 A/Experimental source: strain AR39, HL cells
 C/Genetics: act 1; CP0408
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 60.8%; Score 313; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 3.5e-302;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKEEPFGKLRSEFLPIHTEHKKVLPMLMEFCITFNVTVDRTKTLIVAGPSSGA 60
 DB 1 MTKEEPFGKLRSEFLPIHTEHKKVLPMLMEFCITFNVTVDRTKTLIVAGPSSGA 60
 QY 61 EAIPFIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPFLIFPALTIVIPLRDVL 120
 DB 61 EAIPFIKFWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPFLIFPALTIVIPLRDVL 120
 QY 121 HPTFARLQAITSPGLGLVAIIRNMPFAFYVLAELMGVMSLMFWGPNITITHE 180
 DB 121 HPTFARLQAITSPGLGLVAIIRNMPFAFYVLAELMGVMSLMFWGPNITITHE 180
 QY 181 AKRYALFGICANISLASGRAIYASGLRASVEGVDPMGISRLMAMTIVSGVLMA 240
 DB 181 AKRYALFGICANISLASGRAIYASGLRASVEGVDPMGISRLMAMTIVSGVLMA 240
 QY 241 SYWINKNVLTDPPFYNEEMQKKGAKPRNNKDSFLYIDRSPYITLLTLVIAVIGIC 300
 DB 241 SYWINKNVLTDPPFYNEEMQKKGAKPRNNKDSFLYIDRSPYITLLTLVIAVIGIC 300
 QY 301 INLEIYWKSQLQKQYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKPGWLTALV 360
 DB 301 INLEIYWKSQLQKQYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKPGWLTALV 360
 QY 361 TPVWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTKVALFPSTK 420
 DB 361 TPVWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKTKVALFPSTK 420
 QY 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVICSGIGANTPYLAVILFTI 480
 DB 421 EMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVICSGIGANTPYLAVILFTI 480
 QY 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515
 DB 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515

RESULT 2

ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138) E86534
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C/Accession: E86534
 R/Shirai, M.; Hixkawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ist Nucleic Acids Res. 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A/Reference number: A86491; MUID:20330349; PMID:10871362
 A/Accession: E86534
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-515 <STO>
 A/Cross-references: GB:BA000008; NID:g8978723; PIDN:BAAG8559.1; GSPDB:GN00142
 A/Experimental source: strain J138
 C/Genetics: act 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 43.5%; Score 224; DB 2; Length 515;
 Best Local Similarity 100.0%; Pred. No. 7.9e-214;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 LTVIAYGICINLIEVTKWSQLQYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRK 351
 DB 292 LTVIAYGICINLIEVTKWSQLQYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRK 351
 QY 352 FGMVLTGALVTPVWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411
 DB 352 FGMVLTGALVTPVWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKST 411
 QY 412 KYALPDSKEMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVICSGIGANTPY 471
 DB 412 KYALPDSKEMAYIPLDQEQKVGKKAIDVVAARFGKSGGALLIQGLLVICSGIGANTPY 471
 QY 472 LAVILFTIWIWVSATKINKFLAQSALKEQVADSDAPASS 515
 DB 472 LAVILFTIWIWVSATKINKFLAQSALKEQVADSDAPASS 515

RESULT 3

probable adp/ATP translocase - Chlamydia trachomatis (serotype D, strain UM3/Cx) C71561
 C/Species: Chlamydia trachomatis
 C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C/Accession: C71561
 R/Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract
 A/Reference number: A71570; MUID:99000809; PMID:9784136
 A/Accession: C71561
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-528 <ARN>
 A/Cross-references: GB:AE001281; GB:AE001273; NID:g3328454; PIDN:AA67656.1; PID:g332845;
 A/Experimental source: serotype D, strain UM-3/Cx
 C/Genetics: act 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query March 6.4%; Score 33; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 3.4e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 TKEMAYIPLDQEQKVGKKAIDVVAARFGKSGG 451
 DB 417 TKEMAYIPLDQEQKVGKKAIDVVAARFGKSGG 449

RESULT 4

ADP, ATP carrier protein TC0335 [imported] - Chlamydia muridarum (strain Nig9) C81714

C/Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C/Accession: G81714
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A/Reference number: A81500; MUID:20150255; PMID:10684935
 A/Accession: G81714
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-529 <TER>
 A/Cross-references: GB:AE002301; GB:AE002160; NID:g7190372; PIDN:AAF39198.1; PID:g719037
 A/Experimental source: strain N19g (Mopn)
 C/Genetics:
 A/Gene: TC0335
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 6.4%; Score 33; DB 2; Length 529;
 Best Local Similarity 100.0%; Pred. No. 3,4e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 TKEMAYIPLDGEQKVGKKAIDVVAARFGKSG 451
 Db 417 TKEMAYIPLDGEQKVGKKAIDVVAARFGKSG 449

RESULT 5
 B71707
 ADP/ATP carrier protein (tlc3) RP477 - Rickettsia prowazekii
 C/Species: Rickettsia prowazekii
 C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C/Accession: B71707
 R/Anderson, S.G.E.; Zomrodipour, A.; Anderson, J.O.; Scheritz-Ponten, T.; Almark, U. Nature 396, 133-140, 1998
 A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A/Reference number: A71630; MUID:99039499; PMID:9823893
 A/Accession: B71707
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-501 <AND>
 A/Cross-references: GB:AU235271; GB:AU235269; NID:g3868717; PIDN:CAA14932.1; PID:g386103
 A/Experimental source: strain Madrid E
 C/Genetics:
 A/Gene: tlc3; RP477
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.5%; Score 13; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 LFDSTKEMAYIPL 427
 Db 415 LFDSTKEMAYIPL 427

RESULT 6
 B97790
 ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
 C/Species: Rickettsia conorii
 C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C/Accession: B97790
 R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R. Science 293, 2093-2098, 2001
 A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A/Reference number: A97700; MUID:21442074; PMID:11557893
 A/Accession: B97790
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-501 <KUR>
 A/Cross-references: GB:AE006914; PIDN:AAU03260.1; PID:g15619815; GSPDB:GN00173
 C/Genetics:
 A/Gene: tlc3

C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.5%; Score 13; DB 2; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 LFDSTKEMAYIPL 427
 Db 415 LFDSTKEMAYIPL 427

RESULT 7
 S68205
 ATP/ADP translocase ATP1 precursor - Arabidopsis thaliana
 N/Alternate names: adenine nucleotide translocase; ATP/ADP translocator protein
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 22-Jun-1999
 C/Accession: S68205
 R/Kampfenkel, K.; Mochmann, T.; Batz, O.; van Montagu, M.; Inze, D.; Neuhaus, H.E. FEBS Lett. 374, 351-355, 1995
 A/Title: Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel put
 A/Reference number: S68205; MUID:96059943; PMID:7589569
 A/Accession: S68205
 A/Molecule type: mRNA
 A/Residues: 1-589 <KAM>
 A/Cross-references: EMBL:Z49227; NID:g1051108; PIDN:CAA89201.1; PID:g1051109
 C/Genetics:
 A/Gene: ATP1
 A/Genome: nuclear
 C/Superfamily: rickettsial-type ATP/ADP translocase
 C/Keywords: chloroplast; transmembrane protein
 F:1-100/Domain: transit peptide (chloroplast) #status predicted <TNP>
 F:101-589/Product: ATP/ADP translocase ATP1 #status predicted <MAT>

Query Match 2.1%; Score 11; DB 1; Length 589;
 Best Local Similarity 100.0%; Pred. No. 0.026;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVGKKAIDIV 442
 Db 509 KVGKKAIDIV 519

RESULT 8
 G86288
 Probable adenine nucleotide translocase [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C/Accession: G86288
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chun, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hutzar, L. Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.-C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: G86288
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-618 <STO>
 A/Cross-references: GB:AE005172; NID:g8072388; PIDN:AAF71976.1; GSPDB:GN00141
 C/Genetics:
 A/Map position: 1
 C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.1%; Score 11; DB 2; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.027;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 432 KVKGAALDVV 442
|||||
Db 509 KVKGAALDVV 519

RESULT 9

adenine nucleotide translocase, 19474-21800 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C/Accession: E96834
R/Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltl, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: E96834
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-624 <STO>
A/Cross-references: GB:AE005173; NID:96751705; PIDN:AAE27687.1; GSPDB:GN00141
C/Genetics:
A/Gene: F516.5
A/Map position: 1
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 2.1%; Score 11; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 432 KVKGAALDVV 442
|||||
Db 512 KVKGAALDVV 522

RESULT 10

ATP/ADP-transporter, chloroplast - potato
C/Species: Solanum tuberosum (potato)
C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
C/Accession: T07420
R/Taden, J.; Mohlmann, T.; Kampfenkel, K.; Henrichs, G.; Neuhaus, H.E.
plant J. 16, 531-540, 1998
A/Title: Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuber-
A/Reference number: Z16025
A/Accession: T07420
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-631 <TJA>
A/Cross-references: EMBL:Y10821; PIDN:CAA71785.1
A/Experimental source: cv. Desiree
C/Genetics:
A/Genome: nuclear
C/Superfamily: rickettsial-type ATP/ADP translocase
C/Keywords: chloroplast; membrane protein

Query Match 2.1%; Score 11; DB 2; Length 631;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 432 KVKGAALDVV 442
|||||
Db 510 KVKGAALDVV 520

RESULT 11

B97783

ADP, ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)

C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: B97783
R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro-
Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: B97783
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-511 <KUR>
A/Cross-references: GB:AE006914; PIDN:ALU03204.1; PID:G15619755; GSPDB:GN00173
C/Genetics:
A/Gene: tlc4
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 1.9%; Score 10; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 LFDSTKEMAY 424
|||||
Db 422 LFDSTKEMAY 431

RESULT 12

ADP, ATP carrier protein (tlc4) RP500 - Rickettsia prowazekii
C/Species: Rickettsia prowazekii
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C/Accession: F71653
R/Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A/Reference number: A71630; MUID:99039499; PMID:9823893
A/Accession: F71653
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-512 <AND>
A/Cross-references: GB:AJ235272; GB:AJ235269; NID:93861033; PIDN:CAA14952.1; PID:G386105.
A/Experimental source: strain Madrid E
C/Genetics:
A/Gene: tlc4; RP500
C/Superfamily: rickettsial-type ATP/ADP translocase

Query Match 1.9%; Score 10; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 415 LFDSTKEMAY 424
|||||
Db 422 LFDSTKEMAY 431

RESULT 13

hypothetical protein Rv2529 - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: E70657
R/Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: E70657
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-463 <COL>
A/Cross-references: GB:Z83863; GB:AL123456; NID:93261685; PIDN:CAB06179.1; PID:G1781073

A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV2529
 C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV2529

Query Match
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 AAIIDVVAAR 445
 DB 218 AAIIDVVAAR 226

RESULT 14
 ATP/ADP translocase tlc1 - Rickettsia prowazekii
 N:Alternate names: ADP/ATP carrier protein tlc1; RP053
 C:Species: Rickettsia prowazekii
 C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 03-Nov-2000
 C:Accession: JQ0026; E71713
 R:Williamson, L.R.; Plano, G.V.; Winkler, H.H.; Krause, D.C.; Wood, D.O.
 Gene 80, 269-278, 1989
 A:Title: Nucleotide sequence of the Rickettsia prowazekii ATP/ADP translocase-encoding gene
 A:Reference number: JQ0026; MUID:9060776; PMID:2555259
 A:Accession: JQ0026
 A:Molecule type: DNA
 A:Residues: 1-498 <MIL>
 A:Cross-references: GB:M28916; NID:g152469; PIDN:AAA26382.1; PID:g152470
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sticheritz-Ponten, T.; Alsmark, U.
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:93039499; PMID:9823893
 A:Accession: E71713
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-498 <AND>
 A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14524.1; PID:g386062
 A:Experimental source: strain Madrid E
 C:Comment: This ATP/ADP translocase shares no detectable amino acid sequence homologies
 C:Genetics:
 A:Gene: tlc1; RP053
 C:Function:
 A:Description: exchanges ADP in the rickettsial cell with ATP in the host cell
 C:Superfamily: rickettsial-type ATP/ADP translocase
 C:Keywords: transmembrane protein
 F:28-45/Domain: transmembrane #status predicted <TR01>
 F:68-84/Domain: transmembrane #status predicted <TR02>
 F:93-115/Domain: transmembrane #status predicted <TR03>
 F:183-206/Domain: transmembrane #status predicted <TR04>
 F:219-237/Domain: transmembrane #status predicted <TR05>
 F:279-297/Domain: transmembrane #status predicted <TR06>
 F:346-370/Domain: transmembrane #status predicted <TR07>
 F:380-399/Domain: transmembrane #status predicted <TR08>
 F:445-461/Domain: transmembrane #status predicted <TR09>
 F:466-482/Domain: transmembrane #status predicted <TR10>

Query Match
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 NKXVLTDP 254
 DB 241 NKXVLTDP 249

RESULT 15
 A97710
 ADP/ATP carrier protein [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C:Accession: A97710
 R:Gogate, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Rd

Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:21442074; PMID:11557893
 A:Accession: A97710
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-498 <RUR>
 A:Cross-references: GB:AE006914; PIDN:AAL02619.1; PID:g15619119; GSPDB:GN00173
 C:Genetics:
 A:Gene: tlc1
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 NKXVLTDP 254
 DB 241 NKXVLTDP 249

RESULT 16
 D72723
 hypothetical protein APE0325 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: D72723
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kikuchi, N.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
 A:Reference number: A72450; MUID:9310359; PMID:10382966
 A:Accession: D72723
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-111 <KAM>
 A:Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79280.1; PID:d1043066; PID:g5103
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0325

Query Match
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVGAI 403
 DB 3 LAVVGAI 10

RESULT 17
 E64538
 hypothetical protein HP0149 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: E64538
 R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalil, H.G.; Glodek, A.; McKenney, S.; Venter, A.; Adams, M.D.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: E64538
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-194 <TOM>
 A:Cross-references: GB:AE000536; GB:AE000511; NID:g2313230; PIDN:AAD07223.1; PID:g2313240

Query Match
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 YILLTLTL 293
|||
Db 175 YILLTLTL 182

RESULT 18
D71970
hypothetical protein jhp0137 - Helicobacter pylori (strain J99)
C/Species: Helicobacter pylori
A/Valley: strain J99
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C/Accession: D71970
R/Alt: R.A.; Ling, L.S.L.; Mofit, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A/Reference number: A71800; MUID:99120557; PMID:9923682
A/Accession: D71970
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-197 <ARN>
A/Cross-references: GB:AE001452; GB:AE001439; NID:94154639; PIDN:AD05716.1; PID:9415464
A/Experimental source: strain J99
C/Genetics:
A/Gene: jhp0137

Query Match 1.6%; Score 8; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 YILLTLTL 293
|||
Db 178 YILLTLTL 185

RESULT 19
A53482
transmembrane envelope protein Env - primate T-cell lymphotropic virus type L (fragment)
C/Species: primate T-cell lymphotropic virus type L; PTLV-L
C/Date: 06-Oct-1994 #sequence_revision 25-Apr-1997 #text_change 19-May-2000
C/Accession: A53482
R/Goubau, P.; Van Brussel, M.; Vandamme, A.M.; Liu, H.F.; Desmyter, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 2848-2852, 1994
A/Title: A primate T-lymphotropic virus, PTLV-L, different from human T-lymphotropic vi
A/Reference number: A53482; MUID:94195839; PMID:7908445
A/Accession: A53482
A/Status: preliminary
A/Molecule type: genomic RNA
A/Residues: 1-247 <GCU>
A/Cross-references: GB:Z29673; NID:9474831; PIDN:CA482771.1; PID:9483564
A/Note: sequence extracted from NCBI backbone (NCBIN:146143, NCBI:P.146146)
A/Note: isolate STLV-PH969; host Papio hamadryas
C/Superfamily: type C retrovirus env polyprotein

Query Match 1.6%; Score 8; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 480 IAIWLVA 487
|||
Db 73 IAIWLVA 80

RESULT 20
D72771
probable bacteriophage tail synthase APE0159 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C/Accession: D72771
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamuta, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A/Reference number: A72450; MUID:99310339; PMID:10382966
A/Accession: D72771
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-282 <KAM>
A/Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA79070.1; PID:95103549
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0159
C/Superfamily: Methanococcus jannaschii conserved hypothetical protein M0279

Query Match 1.6%; Score 8; DB 2; Length 282;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 ILLTLTLV 294
|||
Db 234 ILLTLTLV 241

RESULT 21
RMSBNC
T-cell receptor beta chain precursor (F5) - mouse
C/Species: Mus musculus (house mouse)
C/Date: 03-Aug-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
C/Accession: S03716; S25118; A02134; A93336; B93333
R/Palmer, M.S.; Bentley, A.; Gould, K.; Townsend, A.R.M.
Nucleic Acids Res. 17, 2353, 1989
A/Title: The T cell receptor from an influenza-A specific murine CTL clone.
A/Reference number: S03715; MUID:89202046; PMID:2784852
A/Accession: S03716
A/Molecule type: mRNA
A/Residues: 1-307 <PAL>
A/Cross-references: EMBL:X14388; NID:954668; PIDN:CA432563.1; PID:954669
R/Austrup, F.; Kodellja, V.; Kucharzik, T.; Kisch, E.
submitted to the EMBL data library, July 1992
A/Description: Characterization of idiotypic-specific I-Eg-restricted T suppressor lymphoc
/c mice.
A/Reference number: S25117
A/Accession: S25118
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-'G'GT,'6','C','8','V'FV,'14','I','16','N'V,'19','D','21-23','T','25-29','E','31','AE','3'
'K','86','E','88','L','90','S','92','F','94-99','P','101','E','103','K','105-114','LDNSAE'','LY','125-126'
A/Cross-references: EMBL:X67128; NID:954678; PIDN:CA47607.1; PID:954679
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
C/Keywords: glycoprotein; heterotrimer; receptor; T-cell; transmembrane protein
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-307/Product: T-cell receptor beta chain #status predicted <MAT>
F/188-228/Domain: immunoglobulin homology <IMM>
F/281-302/Domain: transmembrane #status predicted <TM>
F/303-307/Domain: intracellular #status predicted <INT>

Query Match 1.6%; Score 8; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 VSGIWLMA 240
|||
Db 293 VSGIWLMA 300

RESULT 22
A57069
olfactory receptor FAT11 - human
C/Species: Homo sapiens (man)
C/Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 26-Aug-1999
C/Accession: A57069
R/Fan, W.; Liu, Y.C.; Parimoo, S.; Weissman, S.M.
Genomics 27, 119-123, 1995
A/Title: Olfactory receptor-like genes are located in the human major histocompatibility
A/Reference number: A57069; MUID:95394447; PMID:765158
A/Accession: A57069

A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-316 <FNA>
 A;Cross-references: GB:J35475; NID:g1041044; PIDN:AA36567.1; PID:g601919
 C;Genetics:
 A;Gene: GDB:FAT11; OLFR2
 A;Cross-references: GDB:1323249; OMIM:600578
 A;Map position: 6p21.3-6p21.3
 C;Superfamily: olfactory receptor OR14

Query Match 1.6%; Score 8; DB 2; Length 316;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
 DB 274 LFYAVGTP 281

RESULT 23
 T02246
 hypothetical protein P1.11659_4 - human
 C;Species: Homo sapiens (man)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 11-Jan-2000
 C;Accession: T02246

R;Lamerdin, J.E.; McCreedy, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankiel submitted to the EMBL Data Library, March 1998
 A;Authors: Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arellano, A.; Description: Sequence analysis of a human P1 clone containing the XRC9 DNA repair gene
 A;Reference number: 214637
 A;Accession: T02246

A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-357 <LMA>
 A;Cross-references: EMBL:AC004472; NID:g2984582; PIDN:AA07983.1; PID:g2984585
 C;Genetics:
 A;Map position: 9
 A;Intons: 2/3; 82/1; 101/3; 135/3; 194/3; 243/1; 269/3; 312/3
 C;Superfamily: erythrocyte band 7 integral membrane protein

Query Match 1.6%; Score 8; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLAAGRA 202
 DB 4 SLAAGRA 11

RESULT 24
 D84646
 hypothetical protein At2g25270 (imported) - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: D84646

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:106117197

A;Accession: D84646
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-360 <STO>
 A;Cross-references: GB:AE002093; NID:g4567250; PIDN:AA223664.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g25270
 A;Map position: 2

Query Match 1.6%; Score 8; DB 2; Length 360;
 Best Local Similarity 100.0%; Pred. No. 16;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 162 VMLSLMF 169
 DB 325 VMLSLMF 332

RESULT 25

KIECA
 acetate kinase (EC 2.7.2.1) - Escherichia coli (strain K-12)

N;Alternate names: acetokine
 C;Species: Escherichia coli

C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 01-Mar-2002
 C;Accession: J00498; J00358; F65001

R;Matsuyama, A.; Yamamoto, H.; Nakano, E.
 J. Bacteriol. 171, 577-580, 1989

A;Title: Cloning, expression, and nucleotide sequence of the Escherichia coli K-12 ackA gene
 A;Reference number: J00498; MUID:89123075; PMID:2536666

A;Accession: J00498
 A;Molecule type: DNA

A;Residues: 1-400 <MAT>
 A;Cross-references: GB:W22956; NID:g145170; PIDN:AAA23406.1; PID:g145171

A;Experimental source: strain K12
 A;Note: parts of the sequence were confirmed by protein sequencing

R;Kakuda, H.; Hosono, K.; Shiroishi, K.; Ichihara, S.
 J. Biochem. 116, 916-922, 1994

A;Title: Identification and characterization of the ackA (acetate kinase A)-pta (phosphotransferase system) operon of Escherichia coli.
 A;Reference number: J00357; MUID:95189796; PMID:7883769

A;Accession: J00358
 A;Molecule type: DNA

A;Residues: 376-400 <KAK>
 A;Cross-references: DDBJ:D17576

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colwell, R.; Rose, D.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.
 A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F65001
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
 A;Residues: 1-400 <BLAT>

A;Cross-references: GB:AE000318; GB:U00096; NID:g1788623; PIDN:AA075356.1; PID:g1788633;
 A;Experimental source: strain K-12, substrain MG1655

C;Comment: Acetate kinase catalyzes the conversion of acetylphosphate to acetate while ADP is converted to ATP. During anaerobic growth of the organism, this enzyme is also involved in the synthesis of acetate.
 C;Genetics:

A;Gene: ackA
 C;Superfamily: acetate kinase

C;Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 1; Length 400;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AARFGSG 450
 DB 363 AARFGSG 370

RESULT 26

D91026
 acetate kinase (imported) - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C;Accession: D91026
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gaezawa, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome comparison with other E. coli strains

A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: D91026

A;Status: preliminary

A/Molecule type: DNA
 A/Residues: 1-400 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA036603.1; PID:G13362650; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RMD 050952
 C/Genetics:
 A/Gene: EC63180
 C/Superfamily: acetate kinase

Query Match 1.6%; Score 8; DB 2; Length 400;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
 |||||
 Db 363 AARFGKSG 370

RESULT 27
 E85870
 acetate kinase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: E85870

R/Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: E85870
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-400 <STO>
 A/Cross-references: GB:AA005174; NID:G12516649; PID:AA057425.1; GSPDB:GN00145; UMGF:Z35
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: ackA
 C/Superfamily: acetate kinase

Query Match 1.6%; Score 8; DB 2; Length 400;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
 |||||
 Db 363 AARFGKSG 370

RESULT 28
 AD0313
 acetate kinase (EC 2.7.2.1) [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C/Accession: AD0313
 R/Perkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360
 A/Accession: AD0313
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-400 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAC91368.1; PID:G15980557; GSPDB:GN00175
 C/Genetics:
 A/Gene: ackA
 C/Superfamily: acetate kinase
 C/Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 2; Length 400;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
 |||||
 Db 363 AARFGKSG 370

RESULT 29
 AG0798
 acetate kinase (EC 2.7.2.1) [imported] - Salmonella enterica subsp. enterica serovar Typh

C/Species: Salmonella enterica subsp. enterica serovar Typh
 A/Note: This species has also been called Salmonella typh
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AG0798

R/Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
 ch, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AG0798

A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-400 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD07569.1; PID:G16503561; GSPDB:GN00176
 C/Genetics:
 A/Gene: STY2567
 C/Superfamily: acetate kinase
 C/Keywords: phosphotransferase

Query Match 1.6%; Score 8; DB 2; Length 400;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 443 AARFGKSG 450
 |||||
 Db 363 AARFGKSG 370

RESULT 30
 S74886
 phycoene dehydrogenase (EC 1.3.-.-) - Synechocystis sp. (strain PCC 6803)

N/Alternate names: phycoene desaturase; protein slr1254
 C/Species: Synechocystis sp.

A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 04-Feb-2000
 C/Accession: S74886; S20939
 R/Kaneko, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S74886
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA

A/Residues: 1-472 <XAN>
 A/Cross-references: EMBL:DB09093; GB:AB001333; NID:G1652844; PID:BA117847.1; PID:dl01858
 A/Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
 R/Martinez-Ferez, I.M.; Vioque, A.
 Plant Mol. Biol. 18, 981-983, 1992

A/Title: Nucleotide sequence of the phycoene desaturase gene from Synechocystis sp. PCC
 A/Reference number: S20939; MUID:92256820; PMID:1581575
 A/Accession: S20939
 A/Molecule type: DNA

A/Residues: 1-451, 'RSPPIST', 459, 'KPF', 463-464, 'PGK' <MAR>
 A/Cross-references: EMBL:X62574
 C/Genetics:
 A/Gene: crtD; pds
 C/Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 1.6%; Score 8; DB 2; Length 472;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 139 GLVALLRN 146
 |||||
 Db 109 GLVALLRN 116

RESULT 31

AD1932
 hypothetical protein all1007 [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp. PCC 7120

A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C/Accession: AD1932

R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Ref. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. PCC 7120

A/Reference number: AB1807; MUID:21595285; PMID:11759840

C/Accession: AD1932

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-480 <KIR>

A/Cross-references: GB:BA000019; PIDN:BA072964.1; PID:g17130353; GSPDB:GN00179

A/Experimental source: strain PCC 7120

C/Genetics:

A/Gene: all1007

Query Match 1.6%; Score 8; DB 2; Length 480;
 Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 288 LLLTLVLI 295
 |||||
 Db 429 LLLTLVLI 436

RESULT 32

ANHU
 angiotensin precursor [validated] - human

N/Alternate names: angiotensinogen

C/Species: Homo sapiens (man)

C/Date: 06-Jul-1992 #sequence_revision 19-Jan-1996 #text_change 08-Dec-2000

C/Accession: A35203; A31362; I37168; A30825; I39462; A90487; A90226; I54281; A01

R/Fukami, A.; Takahashi, S.; Seo, M.S.; Tada, M.; Tanimoto, K.; Uehara, S.; Murakami, J.

Biol. Chem. 265, 7576-7582, 1990

A/Title: Structure and expression of the human angiotensinogen gene. Identification of a

C/Accession: A35203

A/Reference number: A35203; MUID:90237063; PMID:1692023

A/Molecule type: DNA

A/Residues: 1-485 <FKR>

A/Cross-references: GB:X15323; GB:X15324; GB:X15325; GB:X15326; GB:X15327

R/Galliard, I.; Clauser, E.; Corvol, P.

DNA 8, 87-99, 1989

A/Title: Structure of human angiotensinogen gene.

C/Accession: A31362; MUID:89170129; PMID:2924688

A/Molecule type: DNA

A/Residues: 1-267, 'W', 269-332, 'E', 334-485 <GAI>

A/Cross-references: GB:M24686; GB:M24687; GB:M24688

A/Note: the authors translated the codon GAA for residue 333 as Gln

R/Nibu, Y.; Takahashi, S.; Tanimoto, K.; Murakami, K.; Fukami, A.

Biol. Chem. 269, 28598-28605, 1994

A/Title: Identification of cell type-dependent enhancer core element located in the 3'-d

C/Accession: I37168

A/Reference number: I37168; MUID:95050659; PMID:7961807

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-285 <NIB>

A/Cross-references: EMBL:X15324; NID:g1197496; PIDN:CAA33385.1; PID:g1197497

A/Accession: I37169

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 287-375 <NIB>
 A/Cross-references: EMBL:X15325; NID:g28695
 R/Kunapuli, S.P.; Benedict, C.R.; Kumar, A.
 Arch. Biochem. Biophys. 254, 642-646, 1987

A/Title: Tissue specific hormonal regulation of the rat angiotensinogen gene expression.

A/Reference number: A60825; MUID:87212053; PMID:3579332

C/Accession: A60825

A/Molecule type: mRNA

A/Residues: 32-184 <KUN1>

R/Kunapuli, S.P.; Kumar, A.

Circ. Res. 60, 786-790, 1987

A/Title: Molecular cloning of human angiotensinogen cDNA and evidence for the presence o

C/Accession: I39462

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-267, 'W', 269-338 <KUN2>

A/Cross-references: GB:M69110; NID:g178643; PIDN:AAA52282.1; PID:g553181

R/Kageyama, R.; Ohkubo, H.; Nakaneishi, S.

Biochemistry 23, 3603-3609, 1984

A/Title: Primary structure of human preangiotensinogen deduced from the cloned cDNA sequ

A/Reference number: A90487; MUID:8500455; PMID:6089875

C/Accession: A90487

A/Molecule type: mRNA

A/Residues: 1-267, 'W', 269-485 <KAG>

A/Cross-references: GB:K02215; NID:g178639; PIDN:AAA51731.1; PID:g178640

A/Note: it is uncertain whether Met-1 or Met-10 is the initiator

R/Tewksbury, D.A.; Dart, R.A.; Travis, J.

Biochem. Biophys. Res. Commun. 99, 1311-1315, 1981

A/Title: The amino terminal amino acid sequence of human angiotensinogen.

A/Reference number: A90226; MUID:81255848; PMID:7259779

C/Accession: A90226

A/Molecule type: protein

A/Residues: 34-46, 'X', 48-50, 'S', 52-57, 'D' <TEW>

R/Hixson, J.E.; Powers, P.K.

Hum. Genet. 96, 110-112, 1995

A/Title: Detection and characterization of new mutations in the human angiotensinogen ge

A/Reference number: I54281; MUID:95331754; PMID:7607642

C/Accession: I54281

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 176-241, 'I', 243-267, 'W', 269-287, 'ANLSAG' <HIX>

A/Cross-references: GB:S78529; NID:g999316; PIDN:AD14287.1; PID:g4261987

C/Comment: Angiotensin I is released from angiotensinogen by renin, which is secreted in

e I (angiotensin-converting enzyme), primarily in the lungs.

C/Comment: The release of the amino-terminal residue (Asp-34) from angiotensin I and ang

sp-1 angiotensin I is converted to angiotensin III by dipeptidyl carboxypeptidase I.

C/Comment: Angiotensin II causes vasoconstriction by direct action on blood vessels, by

o induces thirst.

C/Comment: Angiotensin II and angiotensin III are equally potent in stimulating the synt

C/Comment: Angiotensinogen is synthesized in the liver and secreted into the plasma. The

C/Genetics:

A/Gene: GDB:AGT

A/Cross-references: GDB:118750; OMIM:106150

A/Map position: 1q42-1q43

A/Intons: 286/1; 375/2; 423/3

C/Superfamily: antithrombin III

C/Keywords: blood pressure control; glycoprotein; liver; plasma; vasoconstrictor

F/1-33/Domain: (or 10-33) signal sequence #status predicted <SIG>

F/34-485/Product: angiotensinogen #status predicted <MPT>

F/34-43/Product: angiotensin I #status experimental <PPI>

F/35-41/Product: angiotensin II #status experimental <PP2>

F/47,170,304,328/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 8; DB 1; Length 485;
 Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 ADRLQAIL 133
 |||||
 Db 155 ADRLQAIL 162

RESULT 33

Transforming protein int-3 - mouse mammary tumor virus

C/Species: mouse mammary tumor virus, MMTV

C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 10-Jul-1998

C/Accession: A38072

R/Robbins, J.; Blondel, B.J.; Gallahan, D.; Callahan, R.

J. Virol. 66, 2594-2599, 1992

A/Title: Mouse mammary tumor gene int-3: a member of the notch gene family transforms ma

A/Reference number: A38072; MUID:92194507; PMID:1312643

A/Accession: A38072

A/Molecule type: mRNA

A/Residues: 1-552 <ROB>

A/Cross-references: GB:M80456

C/Genetics:

A/Gene: int-3

C/Superfamily: int-3 transforming protein; ankyrin repeat homology

C/Keywords: duplication; oncogene; transforming protein

F/217-248/Domain: ankyrin repeat homology <AN1>

F/249-281/Domain: ankyrin repeat homology <AN2>

F/283-315/Domain: ankyrin repeat homology <AN3>

F/316-348/Domain: ankyrin repeat homology <AN4>

Query Match

Best Local Similarity 1.6%; Score 8; DB 1; Length 552;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 390 GTTPLMLA 397

Db 284 GTTPLMLA 291

RESULT 34

Probable large glycine/alanine rich protein - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T36105

R/Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A/Reference number: Z21597

A/Accession: T36105

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1772 <MUR>

A/Cross-references: EMBL:AL049707; PIDN:CA841271.1; GSPDB:GN00070; SCOEDB:SCEL5.02C

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCOEDB:SCEL5.02C

Query Match

Best Local Similarity 1.6%; Score 8; DB 2; Length 1772;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GAGSGGAE 61

Db 1176 GAGSGGAE 1183

RESULT 35

notch4 - mouse

C/Species: Mus musculus (house mouse)

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 08-Sep-2002

C/Accession: T09059

R/Bower, L.; Mahliras, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Sch

submitted to the EMBL Data Library, October 1997

A/Description: Sequence of the mouse major histocompatibility locus class III region.

A/Reference number: Z16543

A/Accession: T09059

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1964 <ROM>

A/Cross-references: EMBL:AF030001; NID:92564945; PID:92564947

C/Genetics:

A/Gene: notch4

A/Map position: 17

A/Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67

1679/3; 1729/1; 1761/3

C/Superfamily: notch protein; ankyrin repeat homology; EGF homology

C/Keywords: receptor; signal transduction

F/514-545/Domain: EGF homology <EGF>

Query Match

Best Local Similarity 1.6%; Score 8; DB 2; Length 1964;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 390 GTTPLMLA 397

Db 1696 GTTPLMLA 1703

RESULT 36

fasciculine-like protein - eastern green mamba

N/Alternate names: protein C1361C1

C/Species: Dendroaspis angusticeps (eastern green mamba)

C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C/Accession: S10464; S21298

R/Ducancel, F.

submitted to the EMBL Data Library, March 1990

A/Reference number: S10464

A/Accession: S10464

A/Molecule type: mRNA

A/Residues: 1-80 <DUC>

A/Cross-references: EMBL:X52260; NID:962757; PIDN:CAA36503.1; PID:962758

R/Ducancel, F.

submitted to the EMBL Data Library, June 1990

A/Reference number: S21298

A/Accession: S21298

A/Molecule type: mRNA

A/Residues: 1-80 <DDW>

A/Cross-references: EMBL:X53410; NID:962753; PIDN:CAA37486.1; PID:962754

C/Superfamily: snake toxin

C/Keywords: toxin

Query Match

Best Local Similarity 1.4%; Score 7; DB 2; Length 80;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLV 294

Db 4 LLLTLV 10

RESULT 37

short toxin 2 precursor - eastern green mamba

N/Alternate names: protein Daf8; toxin F-VIII; toxin TA2

C/Species: Dendroaspis angusticeps (eastern green mamba)

C/Date: 22-Jun-1981 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000

C/Accession: S21299; A01675

R/Ducancel, F.

submitted to the EMBL Data Library, June 1990

A/Reference number: S21299

A/Accession: S21299

A/Molecule type: mRNA

A/Residues: 1-81 <DDC>

A/Cross-references: EMBL:X53409; NID:962755; PIDN:CAA37485.1; PID:962756

U. Biol. Chem. 249, 366-372, 1974

A/Title: Snake venom toxin. The purification and amino acid sequence of toxin TA2 from

A/Reference number: A01675; MUID:74070661; PMID:4203470

A/Accession: A01675

A/Molecule type: protein

A/Residues: 22-53, 'V', 55-81 <VIL>
 C/Comment: Short toxin 2 is similar to toxin 1 in its length and in its nonlethal nature
 and thus forms a fifth type of toxin. Its pharmacological action is unknown. It probab
 C/Superfamily: snake toxin
 C/Keywords: toxin; venom
 F/1-21/Domain: signal sequence #status predicted <SIG>
 F/22-81/Product: short toxin 2 #status experimental <MAT>
 F/24-43,38-60,62-73,74-79/Disulfide bonds: #status predicted

Query Match 1.4%; Score 7; DB 1; Length 81;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLLV 294
 DB 4 LLLTLLV 10

RESULT 38

JCS892

cobrotoxin b - Chinese cobra

C/Species: Naja naja atra (Chinese cobra)

C/Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 21-Jul-2000

C/Accession: JCS892

R/Chang, L.S.; Chou, Y.C.; Lin, S.R.; Wu, B.N.; Lin, J.; Hong, E.; Sun, Y.J.; Hsiao, C.D

J. Biochem. 122, 1252-1259, 1997

A/Title: A novel neurotoxin, cobrotoxin b, from Naja naja atra (Taiwan cobra) venom: Pur

A/Reference number: JCS892; PMID:98158338; PMID:9458573

A/Accession: JCS892

A/Molecule type: DNA

A/Residues: 1-82 <CHA>

A/Cross-references: GB:Y13399; NID:95419941; PIDN:CAV33829.2; PID:95419942

A/Experimental source: venom

C/Genetics:

A/Intons: 19/1; 54/1

C/Superfamily: snake toxin

Query Match 1.4%; Score 7; DB 2; Length 82;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLLV 294
 DB 4 LLLTLLV 10

RESULT 39

NINJ1F

short neurotoxin 1 precursor [validated] - Chinese cobra

N/Alternate names: cobrotoxin 2

C/Species: Naja naja atra (Chinese cobra)

C/Date: 13-Jul-1981 #sequence_revision 17-Mar-2000 #text_change 15-Sep-2000

C/Accession: JCS769; A01698; D40667

R/Chang, L.; Lin, J.; Chou, Y.; Hong, E.

Biochem. Biophys. Res. Commun. 239, 756-762, 1997

A/Title: Genomic structures of cardiotoxin 4 and cobrotoxin from Naja naja atra (Taiwan

A/Reference number: JCS768; PMID:98042466; PMID:9367842

A/Accession: JCS769

A/Molecule type: DNA

A/Residues: 1-83 <CHA>

A/Cross-references: GB:Y12492; NID:95524749; PIDN:CAV33097.2; PID:95524750

A/Experimental source: venom

R/Yang, C.C.; Yang, H.J.; Huang, J.S.

Biochim. Biophys. Acta 188, 65-77, 1969

A/Title: The amino acid sequence of cobrotoxin.

A/Reference number: A90575; PMID:59294410; PMID:5820687

A/Accession: A01698

A/Molecule type: protein

A/Residues: 22-83 <YAN>

R/Chou, S.H.; Raynor, R.L.; Zheng, B.; Chambers, T.C.; Kuo, J.F.

Biochemistry 32, 2062-2067, 1993

A/Title: Cobra venom cardiotoxin (cytotoxin) isoforms and neurotoxin: comparative potenc

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Accession: D40667
 A/Molecule type: protein
 A/Residues: 22-83 <CHI>
 A/Cross-references: PIDN:AA25735.1; PID:9299271

A/Experimental source: venom

A/Note: sequence extracted from NCBI backbone (NCBIP:127034)

R/Yang, C.C.; Yang, H.J.; Chiu, R.H.C.

Biochim. Biophys. Acta 214, 355-363, 1970

A/Title: The position of disulfide bonds in cobrotoxin.

A/Reference number: A90576; PMID:71125117; PMID:5533659

A/Contents: annotation; disulfide bonds

A/Reference number: A52941; PDB:1COD

A/Contents: annotation; conformation by (1)H-NMR, residues 1-62

R/Yu, C.; Bhaskaran, R.; Chuang, L.C.; Yang, C.C.

Biochemistry 32, 2131-2136, 1993

A/Title: Solution conformation of cobrotoxin: a nuclear magnetic resonance and hybrid di

A/Reference number: A58569; PMID:93183875; PMID:8443154

A/Contents: annotation; conformation by (1)H-NMR

C/Genetics:

A/Intons: 19/1; 55/1

C/Superfamily: snake toxin

C/Keywords: neurotoxin; venom

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-83/Product: cobrotoxin #status experimental <MAT>

F/24-45,38-62,64-75,76-81/Disulfide bonds: #status experimental

Query Match 1.4%; Score 7; DB 1; Length 83;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLLV 294
 DB 4 LLLTLLV 10

RESULT 40

NSCH4

nonhistone chromosomal protein HMG-14a - chicken

N/Alternate names: high mobility group protein-14a

C/Species: Gallus gallus (chicken)

C/Date: 30-Jun-1989 #sequence_revision 12-Apr-1996 #text_change 22-Jun-1999

C/Accession: S22122; J02882; J00140

R/Browne, D.L.; Dodgson, J.B.

Submitted to the EMBL Data Library, November 1991

A/Description: The chicken HMG-14a gene is transcribed into multiple mRNAs.

A/Reference number: S22122

A/Accession: S22122

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-105 <BRO>

A/Cross-references: EMBL:X63086

R/Dodgson, J.B.; Browne, D.L.; Black, A.J.

Gene 63, 287-295, 1988

A/Title: Chicken chromosomal protein HMG-14 and HMG-17 cDNA clones: isolation, character

A/Reference number: J02883; PMID:88255874; PMID:3384337

A/Accession: J02882

A/Molecule type: mRNA

A/Residues: 2-105 <DOD>

R/Browne, D.L.; Dodgson, J.B.

Gene 124, 199-206, 1993

A/Title: The gene encoding chicken chromosomal protein HMG-14a is transcribed into multi

A/Reference number: J00140; PMID:93185924; PMID:8444343

A/Accession: J00140

A/Molecule type: DNA

A/Residues: 2-5,16-42 <BR2>

A/Cross-references: EMBL:X63083

C/Comment: The HMG proteins are small nonhistone chromosomal proteins, rich in both basic

C/Genetics: This protein is apparently one of the necessary components of actively transcr

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Accession: A40667

A/Molecule type: protein

A/Residues: 2-5,16-42 <BR2>

A/Cross-references: EMBL:X63083

C/Comment: The HMG proteins are small nonhistone chromosomal proteins, rich in both basic

C/Genetics: This protein is apparently one of the necessary components of actively transcr

A/Reference number: A40667; PMID:93192268; PMID:8448165

A/Accession: A40667

A/Molecule type: protein

A/Residues: 2-5,16-42 <BR2>

A/Cross-references: EMBL:X63083

C/Comment: The HMG proteins are small nonhistone chromosomal proteins, rich in both basic

A: Introns: 5/3; 15/3; 25/3; 42/3; 90/3
C: Superfamily: nonhistone chromosomal protein HMG-17
C: Keywords: chromosomal protein; DNA binding; nucleus

Query Match 1.4%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 263 KGKKGAK 269
|||
Db 59 KGKKGAK 65

Search completed: November 25, 2003, 10:07:35
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 06:57:49 ; Search time 19 Seconds
(without alignments)
1274.671 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 515
Sequence: 1 MTKTEKRPFGKLRSEFLWPIH.....AQSAKKEQVNAQEDSAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	313	60.8	515	1 TLCL1_CHLNP	Q94832 chlamydia p
2	333	6.4	528	1 TLCL1_CHLNP	Q94068 chlamydia t
3	33	6.4	529	1 TLCL1_CHLNP	Q94068 chlamydia m
4	13	2.5	501	1 TLCC1_RICR	Q94067 rickettsia m
5	11	2.1	618	1 TLCC2_ARATH	P92935 arabidopsis
6	11	2.1	624	1 TLCL1_ARATH	Q39002 arabidopsis
7	11	2.1	631	1 TLCL1_SOLTU	Q24381 solanum tub
8	10	1.9	512	1 TLCD1_RICR	Q94067 rickettsia
9	9	1.7	498	1 TLCD1_RICR	P19568 rickettsia
10	8	1.6	173	1 TCBB2_MOUSE	P01851 mus musculu
11	8	1.6	312	1 O2H2_HUMAN	Q95918 homo sapien
12	8	1.6	316	1 O2H1_HUMAN	Q95264 homo sapien
13	8	1.6	316	1 O2H3_HUMAN	Q15062 homo sapien
14	8	1.6	400	1 ACKA_ECOLI	P15046 escherichia
15	8	1.6	400	1 ACKA_SALTY	Q84599 salmonella
16	8	1.6	400	1 ACKA_YERPE	Q84599 yersinia pe
17	8	1.6	472	1 CRTI_SYNT3	P29273 synechocyst
18	8	1.6	485	1 ANGT_HUMAN	P01019 homo sapien
19	8	1.6	1964	1 NTC4_MOUSE	P31695 mus musculu
20	8	1.6	2003	1 NTC4_HUMAN	Q99466 homo sapien
21	7	1.4	35	1 PETG_CVACA	O99199 cyanidium c
22	7	1.4	80	1 TX31_DENAN	P18329 dendroaspis
23	7	1.4	81	1 TXF8_DENAN	P01404 dendroaspis
24	7	1.4	82	1 CBTB_NAUA	P80958 naja atra
25	7	1.4	83	1 CBTB_NAUA	P01430 naja atra
26	7	1.4	104	1 HG15_CHICK	P12902 gallus gall
27	7	1.4	108	1 R33E_SULSO	Q98083 sulfobacill
28	7	1.4	110	1 R33E_METJA	P54061 methanococc
29	7	1.4	146	1 HBB0_PAGBO	P82345 pagotomema
30	7	1.4	146	1 HBB0_PAGBO	P45722 pagotomema
31	7	1.4	146	1 HBB0_PAGBO	P45722 pagotomema
32	7	1.4	168	1 NU6M_LOLAL	O47478 loligo blae
33	7	1.4	172	1 NU6M_CYPCA	P24982 cyprinus ca

34	7	1.4	173	1 NU6M_CARAU	O78689 carassius a
35	7	1.4	173	1 NU6M_GADMO	P55783 gadus morhu
36	7	1.4	185	1 RRP_WIGRO	Q84295 wigleswort
37	7	1.4	191	1 RK24_SPTOL	P27683 spinacia ol
38	7	1.4	192	1 INF_MELCA	P51527 melaeagris g
39	7	1.4	193	1 INF1_CHICK	P42165 gallus gall
40	7	1.4	193	1 INF3_CHICK	O90872 gallus gall
41	7	1.4	194	1 Y537_FUSNN	O84799 fusobacteri
42	7	1.4	236	1 YDUX_ECOLI	P76219 escherichia
43	7	1.4	240	1 HIS4_OCEBH	Q86851 oceanobacil
44	7	1.4	253	1 C151_CERAE	Q94501 ceratophthec
45	7	1.4	253	1 C151_HUMAN	P48509 homo sapien
46	7	1.4	264	1 YMP2_CABEL	O10946 caenorhabdi
47	7	1.4	267	1 SY72_ARATH	Q94665 arabidopsis
48	7	1.4	268	1 NAGE_BORBU	O30564 borrelia bu
49	7	1.4	277	1 SUR4_CABEL	O18864 caenorhabdi
50	7	1.4	295	1 LPLC_BACSU	P39129 bacillus su
51	7	1.4	315	1 COXX_BRUME	O84796 bruceella me
52	7	1.4	322	1 CCSA_CVAPA	P48257 cyanophora
53	7	1.4	331	1 YJFF_ECOLI	P37772 escherichia
54	7	1.4	336	1 GSPK_VIRCH	P45781 vibrio chol
55	7	1.4	340	1 ALP_STRGB	Q94501 streptomyce
56	7	1.4	344	1 OM40_DROME	O94416 drosophila
57	7	1.4	354	1 NU2M_ASTPE	Q33819 aseterina pe
58	7	1.4	375	1 GP27_HUMAN	O94501 homo sapien
59	7	1.4	380	1 CYB_IACUV	O48089 laccera viv
60	7	1.4	380	1 CYB_RANRU	O96623 rana rugosa
61	7	1.4	388	1 PEPF_RABIT	P27823 oryctolagus
62	7	1.4	402	1 YIN2_STRAM	P32426 streptomyce
63	7	1.4	404	1 RSPA_ECOLI	P38104 escherichia
64	7	1.4	442	1 YOH2_BACSU	P54505 bacillus su
65	7	1.4	457	1 NORM_ECOS7	P58164 escherichia
66	7	1.4	457	1 NORM_ECOLI	P77340 escherichia
67	7	1.4	475	1 Z342_HUMAN	Q64404 homo sapien
68	7	1.4	481	1 IMDH_HELPJ	Q92114 helicobacte
69	7	1.4	484	1 IMDH_HELPJ	P56088 helicobacte
70	7	1.4	484	1 Y014_HUMAN	P57764 homo sapien
71	7	1.4	493	1 Y014_HUMAN	O15048 homo sapien
72	7	1.4	516	1 CP11_CAVPO	O06367 cavia porce
73	7	1.4	536	1 FLIF_CAUOR	O04954 caulobacter
74	7	1.4	644	1 YNL5_YEAST	P53925 saccharomyc
75	7	1.4	662	1 SUT2_STYHA	P53392 styloeanthe
76	7	1.4	667	1 SUT1_STYHA	P53391 styloeanthe
77	7	1.4	686	1 ATB2_LISIN	Q92637 listeria in
78	7	1.4	686	1 KLC_STRPU	O05090 strongyloce
79	7	1.4	783	1 YAF3_YEAST	P39719 saccharomyc
80	7	1.4	786	1 TLRI_HUMAN	O15399 homo sapien
81	7	1.4	915	1 Y007_CABEL	O09446 caenorhabdi
82	7	1.4	955	1 AGAB_VIRB7	P48840 vibrio sp.
83	7	1.4	1033	1 ITAB_MOUSE	O94501 mus musculu
84	7	1.4	1137	1 BCSC_ECOS7	Q84501 escherichia
85	7	1.4	1140	1 BCSC_ECOLI	P37650 escherichia
86	7	1.4	1255	1 VGL2_CVISA	P59594 human coron
87	7	1.4	1363	1 XDH_EBENI	O12553 escherichia
88	7	1.4	1429	1 L112_CABEL	P14585 caenorhabdi
89	7	1.4	3011	1 POLG_HCVI	P26664 h genome po
90	7	1.4	3011	1 POLG_HCVI	P27958 h genome po
91	7	1.4	3678	1 DMD_MOUSE	P11531 mus musculu
92	7	1.4	3680	1 DMD_CANFA	O97592 canis famli
93	7	1.4	3685	1 DMD_HUMAN	P11533 homo sapien
94	6	1.2	43	1 DHBD_ASPOR	P80409 aspergillus
95	6	1.2	46	1 RL34_MYCAV	Q92891 mycobacteri
96	6	1.2	48	1 OP01_OYKXI	P83247 oxyopes kit
97	6	1.2	52	1 IPR_SOLME	P01078 solanum mel
98	6	1.2	54	1 LT02_HORVU	O94501 hordeum vul
99	6	1.2	57	1 VSH_MUMPU	P22110 mumps virus
100	6	1.2	58	1 COX5_YEAST	P07255 saccharomyc

RESULT 1

ALIGNMENTS

TLCL CHLTPN STANDARD; PRT; 515 AA.
 AC Q92802; Q92802; Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCL OR ADP 1 OR CPN0351 OR CP0408.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=83358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CM1029;
 RX MEDLINE=99206506; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,
 RA Libber K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eilen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishi K., Hattori M., Kohara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CM1029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001619; AAD18495.1; -;
 CC EMBL; AE002202; AAF38252.1; -;
 CC EMBL; AP002546; BAA98559.1; -;
 CC PIR; E72089; E72089.
 CC PIR; E86534; E86534.
 CC TIGR; CP0408; -;
 CC InterPro; IPR004667; ADP_ATP-car.
 CC Pfam; PF03219; TLC; 1.
 CC TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 266 306 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.
 FT CONFLICT 141 141 V->D (IN REF. 3).
 SQ SEQUENCE 515 AA; 57197 MM; P3P4A52C008E106 CRC64;
 Query Match 60.8%; Score 313; DB 1; Length 515;
 Best Local Similarity 99.6%; Pred. No. 2.56-302;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MTKTEKPRGRLRSEFLPIHTEHLKYLPMFMFCITENTYVLRDRLTIYVAPSGA 60
 DB 1 MTKTEKPRGRLRSEFLPIHTEHLKYLPMFMFCITENTYVLRDRLTIYVAPSGA 60
 QY 61 EAIPIKFWLVPPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFPFIYPLRDVL 120
 DB 61 EAIPIKFWLVPPCAIIFMLIYAKLSNLSKQALFYAVGTPELIFPFIYPLRDVL 120
 QY 121 HPTFADRLQAIPLPGLIGVAILRNWTFPAFYVLAELMGSVMSLMFWGFANITKHE 180
 DB 121 HPTFADRLQAIPLPGLIGVAILRNWTFPAFYVLAELMGSVMSLMFWGFANITKHE 180
 QY 181 AKRYVALFGIGANISLASGAIYVASKLRASVSEGVDPWGISRLIMAMTIVGLVMA 240
 DB 181 AKRYVALFGIGANISLASGAIYVASKLRASVSEGVDPWGISRLIMAMTIVGLVMA 240
 QY 241 SYWINKNVLTDPFPYNEEMQKGAKEPMNMKDSFLYDRSPYILLTLVIAYGIC 300
 DB 241 SYWINKNVLTDPFPYNEEMQKGAKEPMNMKDSFLYDRSPYILLTLVIAYGIC 300
 QY 301 INLEVTWKSQKLQYPMNDYSEFGNFSFWTGVSVLIMLVGQNVIRKRGWLTGALV 360
 DB 301 INLEVTWKSQKLQYPMNDYSEFGNFSFWTGVSVLIMLVGQNVIRKRGWLTGALV 360
 QY 361 TPVVVLLTGIVFPLVYFRNQSGLVAMPGTPTMLLVVGAIONISLSKRYALPSTK 420
 DB 361 TPVVVLLTGIVFPLVYFRNQSGLVAMPGTPTMLLVVGAIONISLSKRYALPSTK 420
 QY 421 EMAYIPLDQEQKVGKAIIVVAFPKSGGALIQGLVYICSGIGAMTPYLAIVLPII 480
 DB 421 EMAYIPLDQEQKVGKAIIVVAFPKSGGALIQGLVYICSGIGAMTPYLAIVLPII 480
 QY 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515
 DB 481 AIWIVSATKINKFLAQSALKEQVADSDAPASS 515
 RESULT 2
 TLCL CHLTPN STANDARD; PRT; 528 AA.
 AC Q94068;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCL OR CT065.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
 NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Ck;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AB001281; AAC67656.1; -
 DR PIR; C71561; C71561.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRPFAM; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SQ SEQUENCE 528 AA; 58117 MW; EAD16A43B95B94B2 CRC64;

Query Match 6.4%; Score 33; DB 1; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1.6e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 419 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 451
 Db 417 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 449

RESULT 3
 TLCL CHLMU STANDARD; PRT; 529 AA.
 AC 09PKX5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADP ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR TC0335.
 OS Chlamydia mirlidrum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_Taxid=83560;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MORN / Ni99;
 RX MEDLINE=20150255; PubMed=10684935; G11 S.R., Heidelberg J.F.,
 RA Read T.D., Brunham R.C., Shen C., Utterback T., Berry K., Baas S.,
 RA White O., Hickey E.K., Peterson J., Utechtick T., Bowman C., Dodson R.,
 RA Linher K., Weidman J., Khouri H., Craven B., Gorman C., Dodson R.,
 RA Gaim M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis MORN and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AB002301; AAF39198.1; -
 DR PIR; C81714; C81714.
 DR TIGR; TC0335; -

DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRPFAM; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 322 342 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SQ SEQUENCE 529 AA; 58304 MW; EED13A4C751071C9 CRC64;

Query Match 6.4%; Score 33; DB 1; Length 529;
 Best Local Similarity 100.0%; Pred. No. 1.6e-24;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 419 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 451
 Db 417 TKEMAYIPLDQEQVKGAIDVVAARFGKSG 449

RESULT 4
 TLCC RICPR STANDARD; PRT; 501 AA.
 AC 09ZDE7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP ATP carrier protein 3 (ADP/ATP translocase 3).
 GN TLCC OR TLG3 OR Rp477.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_Taxid=782;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AJ235271; CAAL4932.1; -
 DR PIR; B71707; B71707.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRPFAM; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.

SQ SEQUENCE 501 AA; 57182 MW; D1EC540E6D7E91F CRC64;
 Query Match 2.5%; Score 13; DB 1; Length 501;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 415 LFDSTKEMAYRPL 427
 DB 415 LFDSTKEMAYRPL 427
 RESULT 5
 TLCL ARATH STANDARD; PRT; 618 AA.
 AC P92935; Q9M9B3;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Chloroplast ADP/ATP carrier protein 2, chloroplast precursor (ADP/ATP translocase 2) (Adenine nucleotide translocase 2).
 GN ATP2 OR AT1G15500 OR T16N11.1.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98206726; PubMed=9546649;
 RA Moehlmann T., Tjaden J., Schweoppe C., Winkler H.H., Kampfenkel K., Neuhaus H.E.;
 RT "Occurrence of two plastidic ATP/ADP transporters in Arabidopsis thaliana L. -- molecular characterisation and comparative structural analysis of similar ATP/ADP translocators from plastids and Rickettsia prowazekii.";
 RT Eur. J. Biochem. 252:353-359 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Columbia;
 RA MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan O., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gall J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luo J.S., Maiti R., Marziani A., Miltchev J., Miranda M., Nguyen M., Niernan W.C., Osborne B.T., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Taiton L.J., Tambunga G., Toriumi M.J., Town C.D., Utechtack T., Van Aken S., Vaysberg M., Vostokala V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
 RT Nature 408:816-820 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Columbia;
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the SGP consortium (Salk/Stanford/PGEC).";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts and other sequence problems.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X94626; CAA64329.1; ALT SEQ.
 DR EMBL: AC013453; AAF71976.1; -
 DR EMBL: AY081350; AAL91239.1; -
 DR EMBL: AY128844; AAM91244.1; -
 DR EMBL: AY084374; AAM60955.1; -
 DR PIR: G86288; G86288.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRfam: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 FT TRANSIT 1
 FT CHAIN 1
 FT CHAIN 1
 SQ SEQUENCE 618 AA; 67530 MW; B2D49E7F17A867A CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 509 KVGKKAIDV 519
 RESULT 6
 TLCL ARATH STANDARD; PRT; 624 AA.
 AC Q39002; Q9AN99; Q9C974;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Chloroplast ADP/ATP carrier protein 1, chloroplast precursor (ADP/ATP translocase 1) (Adenine nucleotide translocase 1).
 GN ATP1 OR AT1G0300 OR F516.5.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Columbia;
 RA MEDLINE=96069943; PubMed=7589569;
 RA Kampfenkel K., Moehlmann T., Batz O., van Montagu M., Inze D., Neuhaus H.E.;
 RT "Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylate translocator of higher plants.";
 RT FEBS Lett. 374:351-355 (1995).
 RN [2]
 RP REVISIONS.
 RA Kampfenkel K.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan O., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gall J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetska I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hopfer S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maltz R., Marzilli A.,
 RA Maltsev J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Patel G., Peterson J., Plam P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Walker M.,
 RA Utechtack T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and Analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RT Nature 408:816-820(2000).
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP FUNCTION: AND SUBCELLULAR LOCATION.
 RX MEDLINE=97117797; PubMed=9025303;
 RA Neuhaus H.E., Thom E., Mohlmann T., Steup M., Kampfenkel K.;
 RT "Characterization of a novel eukaryotic ATP/ADP translocator located
 RT in the plastid envelope of Arabidopsis thaliana L.";
 RL Plant J. 11:73-82(1997).
 CC - FUNCTION: May function as an ATP importer.
 CC - SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC - SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49227; CAAB9201.2; -;
 DR EMBL: AC018848; AAG52434.1; -;
 DR EMBL: AF428316; AAL16246.1; -;
 DR EMBL: AY045903; AAK76577.1; -;
 DR PIR: E96834; E96834.
 DR InterPro: IPR004667; ADP_ATP-car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 DR Transmembrane; Transport; ATP-binding; Multigene family;
 KM Chloroplast; Transit peptide
 FT TRANSLIT 1 ? 624
 FT CHAIN ? 624
 FT TRANSMEM 108 128
 FT TRANSMEM 182 202
 FT TRANSMEM 240 260
 FT TRANSMEM 315 335
 FT TRANSMEM 446 466
 FT TRANSMEM 545 565
 FT CONFLICT 45 94
 FT CONFLICT 103 103
 FT CONFLICT 197 197
 FT CONFLICT 281 281
 SQ SEQUENCE 624 AA; 68134 MW; 530302A89AD08745 CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 512 KVGKKAIDV 522

RESULT 7
 ID TLCD_SOLUT STANDARD; PRT; 631 AA.
 AC 024381;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Plastidic ATP/ADP-transporter.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RA Neuhaus B.;
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC - SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10821; CAA71785.1; -;
 DR PIR: T07420; T07420.
 DR InterPro: IPR004667; ADP_ATP-car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KM Transmembrane; Transport; ATP-binding; Multigene family;
 KM Chloroplast.
 SQ SEQUENCE 631 AA; 68886 MW; 33158380DE29D1EA CRC64;
 Query Match 2.1%; Score 11; DB 1; Length 631;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 432 KVGKKAIDV 442
 DB 510 KVGKKAIDV 520
 RESULT 8
 ID TLCD_RICPR STANDARD; PRT; 512 AA.
 AC 092D47;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 4 (ADP/ATP translocase 4).
 GN TLCD OR TLC4 OR RP500.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Almark U.C.M., Podowski R.M., Naeelund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurilad C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RT Nature 396:133-140(1998).
 CC - FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.

THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLUCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ235272; CAA14952.1; -
 DR PIR: F71653; F71653; ADP_ATP_car.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KM Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 SQ SEQUENCE 512 AA; 56208 MW; D09A265DC38936E8 CRC64;
 Query Match 1.9%; Score 10; DB 1; Length 512;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 415 LFDSTKEMAY 424
 DB 422 LFDSTKEMAY 431
 RESULT 9
 TLCA_RICPR STANDARD; PRT; 498 AA.
 ID TLCA_RICPR STANDARD; PRT; 498 AA.
 AC P19568;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP_ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR TLC1 OR TLC OR RP053.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 NC NCB1_Taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=90060776; PubMed=2555259;
 RA Williams L.R., Plano G.V., Winkler H.H., Krause D.C., Wood D.O.;
 RT "Nucleotide sequence of the Rickettsia prowazekii ATP/ADP translocase-
 RT encoding gene.";
 RL Gene 80:269-278(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Stenicher-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [3]
 RP TOPOLOGY
 RX MEDLINE=91258319; PubMed=1904433;
 RA Plano G.V., Winkler H.H.;
 RT "Identification and initial topological analysis of the Rickettsia
 RT prowazekii ATP/ADP translocase.";
 RL J. Bacteriol. 173:3389-3396(1991).
 RN [4]
 RP TOPOLOGY
 RX MEDLINE=90334349; PubMed=2165754;
 RA Plano G.V., Wood D.O., Winkler H.H.;

"Rickettsia prowazekii and ATP/ADP translocase. Analysis of gene
 RT fusions encoding beta-galactosidase-ATP/ADP translocase fusion
 RT proteins.";
 RL Ann. N.Y. Acad. Sci. 590:397-407(1990).
 RN [5]
 RP TOPOLOGY
 RX MEDLINE=99141397; PubMed=9917392;
 RA Alexeyev M.F., Winkler H.H.;
 RT "Membrane topology of the Rickettsia prowazekii ATP/ADP translocase
 RT revealed by novel dual pho-lac reporters.";
 RL J. Mol. Biol. 285:1503-1513(1999).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC RICKETTSIAL LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLUCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M28816; AAA26382.1; -
 DR EMBL: AJ235270; CAA14524.1; -
 DR PIR: JQ0026; JQ0026.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KM Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 SQ SEQUENCE 498 AA; 56821 MW; D3C450D2BC0BE53E CRC64;
 Query Match 1.7%; Score 9; DB 1; Length 498;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 246 NKNVLTDP 254
 DB 241 NKNVLTDP 249
 RESULT 10

TCB2_MOUSE
ID TCB2_MOUSE STANDARD; PRT; 173 AA.
AC P01851;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE T-cell receptor beta-2 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OK NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A;
RX MEDLINE=84270704; PubMed=6336329;
RA Gascogne N.R.J., Chien Y., Becker D.M., Kavalier J., Davis M.M.;
RT "Genomic organization and sequence of T-cell receptor beta-chain
constant- and joining-region genes."
RL Nature 310:387-391(1984).
RN [2]
RP SEQUENCE FROM N.A. (CLONE 2C).
RC STRAIN=BALB.B;
RX MEDLINE=84245824; PubMed=6330561;
RA Saito H., Kranz D.M., Takagaki Y., Hayday A.C., Eisen H.N.,
RA Tonegawa S.;
RT "Complete primary structure of a heterodimeric T-cell receptor
deduced from cDNA sequences."
RL Nature 309:757-762(1984).
CC -1- MISCELLANEOUS: CLONE B10.A WAS ISOLATED FROM A CYTOTOXIC T
LYMPHOCYTE.
CC -1- MISCELLANEOUS: CLONE 2C WAS ISOLATED FROM A CYTOTOXIC T
LYMPHOCYTE.
CC HSPSP; P01842; 2MCG.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG_1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW T-cell; Receptor; Immunoglobulin domain; Transmembrane; Glycoprotein.
FT T-CELL 1
FT NON_TER 1
FT DOMAIN 1 146 C REGION.
FT TRANSMEM 147 168
FT DOMAIN 169 173 CYTOPLASMIC TAIL.
FT VARIANT 50 50 K->R (IN CLONE 2C).
FT VARIANT 70 70 Y->H (IN CLONE 2C).
SQ SEQUENCE 173 AA; 19297 MW; A5458149614CF295 CRC64;
Query Match 1.6%; Score 8; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 VSGGLVLA 240
DB 159 VSGGLVLA 166
RESULT 11
ID 02H2_HUMAN STANDARD; PRT; 312 AA.
AC O95918;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Olfactory receptor 2H2 (Hs6M1-12).
GN OR2H2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031983; CAA21455.1; -
DR InterPro; IPR00276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_Rhodopsn.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 23
FT TRANSMEM 24 47 1 (POTENTIAL).
FT DOMAIN 48 55 2 (POTENTIAL).
FT TRANSMEM 56 77 2 (POTENTIAL).
FT DOMAIN 78 98 3 (POTENTIAL).
FT TRANSMEM 99 118 3 (POTENTIAL).
FT DOMAIN 119 137 4 (POTENTIAL).
FT TRANSMEM 138 156 4 (POTENTIAL).
FT DOMAIN 157 193 5 (POTENTIAL).
FT TRANSMEM 194 217 5 (POTENTIAL).
FT DOMAIN 218 234 6 (POTENTIAL).
FT TRANSMEM 235 257 6 (POTENTIAL).
FT DOMAIN 258 270 7 (POTENTIAL).
FT TRANSMEM 271 290 7 (POTENTIAL).
FT DOMAIN 291 312 7 (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.
FT CARBOHYD 3 3 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 312 AA; 34797 MW; 2C7AB6C7E08ABBB CRC64;
Query Match 1.6%; Score 8; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281
RESULT 12
ID 02H1_HUMAN STANDARD; PRT; 316 AA.
AC Q9GZK4; Q43629; Q43661; Q9GZK9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Olfactory receptor 2H1 (Hs6M1-16) (Olfactory receptor 6-2) (OR6-2)
GN OR2H1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ASN-63.
RA Ziegler A., Ehlers A., Forbes S.A., Trowsdale J., Uchanska-Ziegler B.,
RA Voiz A., Younger R., Beck S.;
RT "Polymorphic olfactory receptor genes and HLA loci constitute extended
haploypes."
RL (in) Kasahara M. (eds.);
RL Major histocompatibility complex-evolution, structure, and function,
RL pp.110-130, Springer-Verlag, Tokyo (2000).
RN [2]

RP SEQUENCE FROM N.A.
RC Tissue-Testis:
RA Volz A., Ehlers A., Younger R., Forbes S., Trowdale J., Beck S.,
RA Ziegler A.;
RT "Complex transcriptional control of MHC-linked olfactory receptor
RT genes includes long distance and extensive alternative splicing, exon
RT splicing and premature polyadenylation.";
RT Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Whitaker H.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 58-251 FROM N.A.
RA Galliano H.;
RT "Olfactory receptor gene cluster in man and mouse major
RT histocompatibility complex.";
RL Immunogenetics 0:0-0(1998).
RN [5]
RP SEQUENCE OF 58-251 FROM N.A.
RA Amadou C., Avouet P., Ribouchon M.-T., Bouissou C., Tazi-Ahmini R.,
RA Ayer C., Pontarotti P.;
RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ302604; CAC20524.1; -
DR EMBL; AJ302605; CAC20525.1; -
DR EMBL; AJ302606; CAC20526.1; -
DR EMBL; AJ302607; CAC20527.1; -
DR EMBL; AJ302608; CAC20528.1; -
DR EMBL; AJ302609; CAC20529.1; -
DR EMBL; AJ302610; CAC20530.1; -
DR EMBL; AJ302611; CAC20531.1; -
DR EMBL; AJ302612; CAC20532.1; -
DR EMBL; AJ302613; CAC20533.1; -
DR EMBL; AJ459847; CAD31038.1; -
DR EMBL; AL035542; CAB44506.1; -
DR EMBL; AF042078; AAC00184.1; -
DR EMBL; AF044491; AAC00188.1; -
DR GeneW; HGNC:8252; OR2H1.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_REC_P2_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction; Polymorphism.
FT VARIANT 63 63 D->N (IN ALLELE 6M1-16-02).
FT FT
FT CONFLICT 158 158 /FTID=VAR_010944.
FT CONFLICT 184 184 T->K (IN REF. 5).
FT CONFLICT 215 215 R->G (IN REF. 5).
FT CONFLICT 215 215 S->Y (IN REF. 5).
SQ SEQUENCE 316 AA; 35338 MW; 19140A7FD690CF85 CRC64;
Query Match 1.6%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
ID 02H3 HUMAN STANDARD; PRT; 316 AA.
AC Q15062;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Olfactory receptor 2H3 (Olfactory receptor-like protein FAT11).
GN OR2H3 OR OLFR2 OR FAT11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394447; PubMed=7665158;
RA Fan W., Liu Y.-C., Paximou S., Weissman S.M.;
RT "Olfactory receptor-like genes are located in the human major
RT histocompatibility complex.";
RL Genomics 27:119-123(1995).
RN [2]
RP SEQUENCE OF 10-312 FROM N.A., AND VARIANTS SER-30; VAL-48 AND ALA-220.
RX MEDLINE=20341390; PubMed=10880742;
RA Eklund A.C., Belchak M.M., Lapidos K., Raha-Chowdhury R., Ober C.;
RT "Polymorphisms in the HLA-linked olfactory receptor genes in the
RT Hutterites.";
RL Hum. Immunol. 61:711-717(2000).
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35475; AAB36567.1; -
DR EMBL; AF211939; AAF98751.1; -
DR EMBL; AF211940; AAF98752.1; -
DR EMBL; AF211941; AAF98753.1; -
DR EMBL; AF211942; AAF98754.1; -
DR PIR; A57069; A57069.
DR GeneW; HGNC:8254; OR2H3.
DR MIM; 600578; -
DR GO; GO:0003793; F:defense/immunity protein activity; TAS.
DR GO; GO:004984; F:olfactory receptor activity; TAS.
DR GO; GO:007618; P:matng; TAS.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_REC_P2_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 23 46 1 (POTENTIAL).
FT DOMAIN 47 55 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 56 77 2 (POTENTIAL).
FT DOMAIN 78 98 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 99 118 3 (POTENTIAL).
FT DOMAIN 119 137 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 138 158 4 (POTENTIAL).
FT DOMAIN 159 198 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 199 220 5 (POTENTIAL).
FT DOMAIN 221 234 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 235 259 6 (POTENTIAL).
FT DOMAIN 260 270 7 (POTENTIAL).
FT TRANSSEM 271 290 7 (POTENTIAL).
FT DOMAIN 291 316 CYTOPLASMIC (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.

FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 30 30 F->S.
 FT VARIANT 48 48 /FTId=VAR_010227.
 FT VARIANT 220 220 A->V.
 FT VARIANT 220 220 T->A.
 FT CONFLICT 22 22 E->G (IN REF. 1).
 FT CONFLICT 28 28 V->D (IN REF. 1).
 FT CONFLICT 30 30 F->I (IN REF. 1).
 FT CONFLICT 51 51 P->T (IN REF. 1).
 FT CONFLICT 81 81 V->A (IN REF. 1).
 FT CONFLICT 113 114 LT->MK (IN REF. 1).
 FT CONFLICT 153 153 E->G (IN REF. 1).
 FT CONFLICT 230 232 KGR->TAM (IN REF. 1).
 FT CONFLICT 267 267 E->G (IN REF. 1).
 FT CONFLICT 287 287 I->V (IN REF. 1).
 FT CONFLICT 295 296 VT->IK (IN REF. 1).
 FT CONFLICT 299 299 F->L (IN REF. 1).
 FT CONFLICT 307 311 RDSRE->MGITQ (IN REF. 2).
 SQ SEQUENCE 316 AA; 35410 MM; BE298672B663A7CE CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 316;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 DB 274 LFYAVGTP 281
 RESULT 14
 ACRA_ECOLI STANDARD; PRT; 400 AA.
 ID ACRA_ECOLI
 AC P1504; P78188; Q59386;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Acetate kinase (EC 2.7.2.1) (acetate kinase).
 GN ACRA OR ACK OR B2296 OR C2838 OR Z3558 OR ECS3180.
 OS Escherichia coli.
 OS Escherichia coli O6, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=K12;
 RX MEDLINE=89123075; PubMed=2536666;
 RA Matsuyama A., Yamamoto H., Nakano E.;
 RT "Cloning, expression, and nucleotide sequence of the Escherichia coli
 RT K-12 ackA gene.";
 RL J. Bacteriol. 171:577-580(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / KH131;
 RX MEDLINE=95189796; PubMed=7883769;
 RA Kakuda H., Hosono K., Shirotani K., Ichihara S.;
 RT "Identification and characterization of the ackA (acetate kinase
 RT A)-pta (phosphotransacetylase) operon and complementation analysis of
 RT acetate utilization by an ackA-pta deletion mutant of Escherichia
 RT coli.";
 RL J. Biochem. 116:916-922(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayaishi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Saito Y., Sivasubraman S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horinouchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett U., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett U., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimmalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION OF ACETATE TO ACETYL COA AND
 CC THE SECRETION OF ACETATE. DURING ANAEROBIC GROWTH OF THE ORGANISM,
 CC THIS ENZYME IS ALSO INVOLVED IN THE SYNTHESIS OF MOST OF THE ATP
 CC FORMED CATABOLICALLY.
 CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
 CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the acetate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: M22956; AAA23406.1; -;
 DR EMBL: D17576; BAA04501.1; -;
 DR EMBL: AB000318; AAC75356.1; -;
 DR EMBL: D90861; BAA16135.1; -;
 DR EMBL: D90860; BAA16131.1; -;
 DR EMBL: AB016763; AAN81292.1; -;


```

DR EMBL; AB005461; AAC57425.1; -
DR EMBL; AP002561; BAB36603.1; -
DR PIR; D91026; D91026.
DR PIR; J0498; KIECAA.
DR PDB; 1LRG; 19-JUN-02.
DR SWISS-2DPAGE; P15046; COLI.
DR ECODBASE; G041.3; 6TH EDITION.
DR Ecocore; E010027; ackA.
DR HAMAP; MF_00020; -; 1.
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR004372; AckA.
DR Pfam; PF00871; Acetate_kinase; 1.
DR PRINTS; PR00471; ACETATEKINASE.
DR TIGRFAMs; TIGR00016; ackA; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
DR TRANSFERASE; Kinase; Complete proteome; 3D-structure.
DR CONFLICT 221 221 R -> G (IN REF. 2).
DR CONFLICT 240 240 M -> I (IN REF. 2).
DR SEQUENCE 400 AA; 43290 MW; 709C855C514656C8 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
Db 363 AARFGKSG 370

RESULT 15
ID ACKA_SALTY STANDARD; PRT; 400 AA.
AC Q8XJ97;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acetate kinase (EC 2.7.2.1) (Acetokinese).
GN ACKA OR STM2337 OR STY2567 OR T0527.
OS Salmonella typhimurium, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601.
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=Lt2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portolick S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RT Nature 413:852-856(2001).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Pakhilli J., Dougan G., James K.D., Thomson N.R., Pichard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Baham D., Brooks K., Chillingworth T., Compton P.,
RA Cronin A., Davis P., Davies R.W., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hen T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leazer S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford S., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhimurium CT18."
RT Nature 413:848-852(2001).
RL [3]
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC SPECIES=S. typhimurium; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanski V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhimurium Ty2
and CT18."
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: Belongs to the acetokinese family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AB008805; AA21238.1; -
DR EMBL; AL627274; CA007569.1; -
DR EMBL; AB016836; AA068233.1; -
DR HAMAP; MF_00020; -; 1.
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR004372; AckA.
DR Pfam; PF00871; Acetate_kinase; 1.
DR PRINTS; PR00471; ACETATEKINASE.
DR TIGRFAMs; TIGR00016; ackA; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
DR TRANSFERASE; Kinase; Complete proteome.
DR SEQUENCE 400 AA; 43257 MW; BD26BC1AB08CA7AD CRC64;

Query Match 1.6%; Score 8; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
Db 363 AARFGKSG 370

RESULT 16
ID ACKA_YERPE STANDARD; PRT; 400 AA.
AC Q8ZDU6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acetate kinase (EC 2.7.2.1) (Acetokinese).
GN ACKA OR YFO2566 OR Y1622.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Pakhilli J., Wren B.W., Thomson N.R.,
RA Baker S., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Bentley S.D., Bentley S.D., Brooks K., Cerdano-Terraza A.M.,
RA Chillingworth T., Cronin A., Davies R.W., Davis F., Dougan G.,
RA Feltham T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leazer S., Moule S., Oyston P.C.F., Quail M., Rutherford S.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague."
RT Nature 413:523-527(2001).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;

```



```

RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Petherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RA "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: Belongs to the acetylkinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ14152; CAC91368.1; -.
CC EMBL; AE013765; AAM85191.1; -.
CC PIR; AD0313; AD0313.
CC HAMAP; MF_00020; -. 1.
CC InterPro; IPR0000890; Acetate_kin.
CC InterPro; IPR004172; ACKA.
CC Pfam; PF00871; Acetate_Kinase; 1.
CC PRINTS; PR00471; Acetate_Kinase.
CC TIGRFAMs; TIGR00016; ACKA; 1.
CC PROSITE; PS01075; ACETATE_KINASE_1; 1.
CC PROSITE; PS01076; ACETATE_KINASE_2; 1.
CC TRANSFAM; PF01593; Amino oxidase; 1.
CC KMW TRANSFAM; Kinase; Complete proteome.
SQ SEQUENCE 400 AA; 43018 MW; 21BA6CC5B1A27716 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AARFGKSG 450
DB 363 AARFGKSG 370

RESULT 17
CRTL SYNY3
ID CRTL SYNY3 STANDARD; PRT; 472 AA.
AC P29273;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN PDS OR CRD OR SLR1254.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92256820; PubMed=1581575;
RA Martinez-Ferez I.M., Vioque A.;
RT "Nucleotide sequence of the phytoene desaturase gene from
RT Synechocystis sp. PCC 6803 and characterization of a new mutation
RT which confers resistance to the herbicide norflurazon.";
RL Plant Mol. Biol. 18:981-983(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Vioque A.;
RL Submitted (NOV-1993) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kocani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K.,

```

```

RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA Yanada M., Yasuda M., Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE
CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO
CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- ENZYME REGULATION: INHIBITED BY THE HERBICIDE NORFLURAZON IN A
CC NON-COMPETITIVE WAY.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (Probable).
CC -1- SIMILARITY: BELONGS TO THE PHYTOENE DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62574; CAA44452.1; -.
CC EMBL; D90809; BAA17847.1; -.
CC PIR; S74886; S74886.
CC InterPro; IPR000759; Adrxn_reductase.
CC InterPro; IPR001613; Aminooxid fl.
CC InterPro; IPR002937; Amino oxidase.
CC Pfam; PF01593; Amino oxidase; 1.
CC PRINTS; PR00419; AMINOXIDASE.
CC PRINTS; PR00757; AMINOXIDASE.
CC Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
KW Membrane; Herbicide resistance; Complete proteome.
FT NP BIND 7 23 FAD (ADP PART) (POTENTIAL).
FT VARIANT 195 195 R->C (CONFERS RESISTANCE TO THE
FT SEQUENCE 472 AA; 52920 MW; 81D089A6DMA28758 CRC64;
HERBICIDE NORFLURAZON).

Query Match 1.6%; Score 8; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GLVAIRN 146
DB 109 GLVAIRN 116

RESULT 18
ANGT HUMAN
ID ANGT HUMAN STANDARD; PRT; 485 AA.
AC P01019; Q16358; Q16359; Q96F91;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Angiotensinogen precursor [Contains: Angiotensin I (Ang I);
DE Angiotensin II (Ang II); Angiotensin III (Ang III) (Des-Asp1)] -
DE Angiotensin II]].
GN AGT OR SERPINB8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89170129; PubMed=2924688;
RA Galliard I., Clauser E., Corvol P.;
RT "Structure of human angiotensinogen gene.";
RL DNA 8:87-99(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85000455; PubMed=6089875;

```

RA Kageyama R., Ohkubo H., Nakanishi S.;
 RT "Primary structure of human preangiotensinogen deduced from the
 RT cloned cDNA sequence.";
 RL Biochemistry 23:3603-3609 (1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90237063; PubMed=1692023;
 RA Fukumizu A., Takahashi S., Seo M.S., Tada M., Tanimoto K., Uehara S.,
 RA Murakami K.;
 RT "Structure and expression of the human angiotensinogen gene.
 RT Identification of a unique and highly active promoter.";
 RL J. Biol. Chem. 265:7576-7582 (1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheaffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stabileton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5]
 RP SEQUENCE OF 1-338 FROM N.A.
 RX MEDLINE=87244745; PubMed=2685106;
 RA Kunapuli S.P., Kumar A.;
 RT "Molecular cloning of human angiotensinogen cDNA and evidence for the
 RT presence of its mRNA in rat heart.";
 RL Circ. Res. 60:786-790 (1987).
 RN [6]
 RP SEQUENCE OF 34-45, AND SUBUNITS.
 RC TISSUE=Serum;
 RX MEDLINE=95293954; PubMed=7539791;
 RA Oxyg C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
 RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
 RT "Identification of angiotensinogen and complement C3dg as novel
 RT proteins binding the proform of eosinophil major basic protein in
 RT human pregnancy serum and plasma.";
 RL J. Biol. Chem. 270:13645-13651 (1995).
 RN [7]
 RP SEQUENCE OF 34-43.
 RX MEDLINE=69014170; PubMed=4300938;
 RA Arkawa K., Minohara A., Yamada J., Nakamura M.;
 RT "Enzymatic degradation and electrophoresis of human angiotensin I.";
 RL Biochim. Biophys. Acta 168:106-112 (1988).
 RN [8]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=86056581; PubMed=3934016;
 RA Campbell D.J., Bounhik J., Coezy E., Menard J., Corvol P.;
 RT "Processing of rat and human angiotensinogen precursors by microsomal
 RT membranes.";
 RL Mol. Cell. Endocrinol. 43:31-40 (1985).
 RN [9]
 RP FUNCTION OF ANGIOTENSIN III.
 RX MEDLINE=7516649; PubMed=1132082;
 RA Goodfriend T.L., Peach M.J.;
 RT "Angiotensin III: (DES-Aspartic Acid-1)-Angiotensin II. Evidence and
 RT speculation for its role as an important agonist in the renin -
 RT angiotensin system.";

RL Circ. Res. 36:38-48 (1975).
 RN [10]
 RP STRUCTURE BY NMR OF ANGIOTENSIN II.
 RX MEDLINE=98151281; PubMed=9492317;
 RA Carpenter K.A., Wilkes B.C., Schiller P.W.;
 RT "The octapeptide angiotensin II adopts a well-defined structure in a
 RT phospholipid environment.";
 RL Eur. J. Biochem. 251:448-453 (1998).
 RN [11]
 RP VARIANTS MET-207, THR-268 AND CYS-281.
 RX MEDLINE=93008239; PubMed=1394429;
 RA Jeunemaitre X., Soubrier F., Kotelevtsev Y.V., Lifton R.P.,
 RA Williams C.S., Charu A., Hunt S.C., Hopkins P.N., Williams R.R.,
 RA Lalouel J.-M., Corvol P.;
 RT "Molecular basis of human hypertension: role of angiotensinogen.";
 RL Cell 71:169-180 (1992).
 RN [12]
 RP VARIANT THR-268.
 RX MEDLINE=93291876; PubMed=8513325;
 RA Ward K., Hata A., Jeunemaitre X., Helin C., Nelson L., Nankawa C.,
 RA Farrington P.F., Ogasawara M., Suzumori K., Tomoda S., Beritibi S.,
 RA Sasaki M., Corvol P., Lifton R.P., Lalouel J.-M.;
 RT "A molecular variant of angiotensinogen associated with
 RT preclampsia.";
 RL Nat. Genet. 4:59-61 (1993).
 RN [13]
 RP VARIANTS ILE-242; ARG-244 AND CYS-281.
 RX MEDLINE=95331754; PubMed=7607642;
 RA Hixson J.E., Powers P.K.;
 RT "Detection and characterization of new mutations in the human
 RT angiotensinogen gene (AGT).";
 RL Hum. Genet. 96:110-112 (1995).
 RN [14]
 RP CHARACTERIZATION OF VARIANT CYS-281.
 RX MEDLINE=96199253; PubMed=862167;
 RA Gimenez-Roqueplo A.P., Lecoq I., Cohen P., Simon D., Guyene T.T.,
 RA Celerier J., Pau B., Corvol P., Clausen E., Jeunemaitre X.;
 RT "The natural mutation Y248C of human angiotensinogen leads to abnormal
 RT glycosylation and altered immunological recognition of the protein.";
 RL J. Biol. Chem. 271:9838-9844 (1996).
 RN [15]
 RP FUNCTION: IN RESPONSE TO LOWERED BLOOD PRESSURE, THE ENZYME RENIN
 CC CLEAVES ANGIOTENSIN I, FROM ANGIOTENSINOGEN. ACE (ANGIOTENSIN
 CC CONVERTING ENZYME) THEN REMOVES A DIPEPTIDE TO YIELD THE
 CC PHYSIOLOGICALLY ACTIVE PEPTIDE ANGIOTENSIN II, THE MOST POTENT
 CC PRESSOR SUBSTANCE KNOWN, WHICH HELPS REGULATE VOLUME AND MINERAL
 CC BALANCE OF BODY FLUIDS.
 CC - FUNCTION: Angiotensin III stimulates aldosterone release.
 CC - SUBUNIT: During pregnancy, exists as a disulfide-linked 2:2
 CC heterotrimer with the proform of PRG2 and as a complex (probably
 CC a 2:2:2 heterotrimer) with pro-PRG2 and C3dg.
 CC - SUBCELLULAR LOCATION: Secreted.
 CC - TISSUE SPECIFICITY: Synthesized by the liver and secreted in the
 CC plasma.
 CC - DISEASE: AGT SEEMS TO BE ASSOCIATED WITH A PREDISPOSITION TO
 CC ESSENTIAL HYPERTENSION AS WELL AS PREGNANCY-INDUCED HYPERTENSION
 CC (PIH) (PRECLAMPSIA).
 CC - SIMILARITY: BELONGS TO THE SERPIN FAMILY.
 CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-10 IS THE INITIATOR.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: K02215; AAA51731.1; -;
 CC EMBL: M24689; AAA51679.1; -;
 CC EMBL: M24686; AAA51679.1; JOINED.
 CC EMBL: M24687; AAA51679.1; JOINED.
 CC EMBL: M24688; AAA51679.1; JOINED.
 CC EMBL: X15324; CAA33385.1; -;

DR EMBL: X15325; CAA33385.1; JOINED.
 DR EMBL: X15326; CAA33385.1; JOINED.
 DR EMBL: X15327; CAA33385.1; JOINED.
 DR EMBL: M69110; AAA52282.1; -.
 DR EMBL: BC011519; AAH11519.1; -.
 DR EMBL: S78529; AAD14287.1; -.
 DR EMBL: S78530; AAD14288.1; -.
 DR PIR: A35203; ANHU.
 DR SWISS-2DPAGE; P01019; HUMAN.
 DR Genew; HNCN333; AGT.
 DR MIM; 106150; -.
 DR GO; GO:0005625; C:soluble fraction; TAS.
 DR GO; GO:0004667; F:serine protease inhibitor activity; TAS.
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
 DR GO; GO:0007677; P:cell-cell signalling; TAS.
 DR GO; GO:0007565; P:pregnancy; TAS.
 DR GO; GO:0008217; P:regulation of blood pressure; TAS.
 DR InterPro; IPR000227; Angiotensin.
 DR Pfam; PPR0079; Serpin.
 DR PRINTS; PR00654; ANGIOTENSIN.
 DR SMART; SM00093; SERPIN; 1.
 DR PROSITE; PS00284; SERPIN; 1.
 KW Vasoconstrictor; Glycoprotein; Plasma; Serpin; Signal;
 KM Disease mutation; Polymorphism.
 FT SIGNAL 1 33
 FT CHAIN 1 485 ANGIOTENSINOGEN.
 FT PEPTIDE 34 43 ANGIOTENSIN I.
 FT PEPTIDE 34 41 ANGIOTENSIN II.
 FT PEPTIDE 35 41 ANGIOTENSIN III.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .).
 Query Match 1.6%; Score 8; DB 1; Length 485;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 ADRLQAIL 133
 DB 155 ADRLQAIL 162

RESULT 19
 NTC4_MOUSE
 ID NTC4_MOUSE STANDARD; PRT; 1964 AA.
 AC P31695; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neogenic locus notch homolog protein 4 precursor (Notch 4)
 GN [contains: Transforming protein Int-3].
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells."; J. Virol. 66:2594-2599(1992).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294559; PubMed=9150355;
 RA Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4)."; Oncogene 14:1883-1890(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Testis;

RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sasseon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene."; Development 122:2251-2259(1996).
 RL [4]
 RN SEQUENCE FROM N.A.
 RP Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region."; Submitted (OCT-1997) to the EMBL/GenBank/DBD databases.
 RL [5]
 RN SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracellular type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events."; J. Virol. 73:5166-5171(1999).
 RL [6]
 RN FUNCTION.
 RP MEDLINE=21244657; PubMed=11344305;
 RX Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium."; Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RL [7]
 RN SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 RP OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent
 RT proteolysis."; J. Biol. Chem. 276:40268-40273(2001).
 RL [8]
 RN POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members."; Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 RL [9]
 RN FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(BC) which are probably linked by disulfide
 CC bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytic processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 dpc.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(BC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.

CC -1- DISEASE: Loss of the extracellular domain causes constitutive
 CC activation of the Notch protein, which leads to hyperproliferation
 CC of glandular epithelial tissues and development of mammary
 CC carcinomas.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: Contains 29 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M80456; AAB38377.1; -.
 CC EMBL; U43691; AAC52631.1; -.
 CC EMBL; U43691; AAC52631.1; -.
 CC EMBL; AF030001; AAB82004.1; -.
 CC EMBL; AB016771; BAA32281.1; ALT SEQ.
 CC EMBL; AB016772; BAA32283.1; ALT_INIT.
 CC EMBL; AB016773; BAA32284.1; ALT_INIT.
 CC EMBL; AB016774; BAA32285.1; -.
 CC PIR; T09059; T09059.
 CC HSSP; P08709; 1BR9.
 CC MGD; MGI:107471; N0cch4.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_IT.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR02049; Laminin_EGF.
 CC InterPro; IPR008000; Notch.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00068; EGF; 27.
 CC Pfam; PF00066; notch; 2.
 CC PRINTS; PR00010; EGFBL00D.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 11.
 CC SMART; SM00004; NL_2.
 CC PROSITE; PS50088; ANK_REPEAT; 5.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS00010; ASX_HYDROXYL; 11.
 CC PROSITE; PS00022; EGF_1; 28.
 CC PROSITE; PS01186; EGF_2; 21.
 CC PROSITE; PS01187; EGF_CA; 9.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 CC SIGNAL
 CC CHAIN 1 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
 CC CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.
 CC CHAIN 1428 1964 NOTCH EXTRACELLULAR TRUNCATION.
 CC CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.
 CC TRANSMEM 1444 1464 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1444 1464 POTENTIAL.
 CC TRANSMEM 1465 1964 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 21 60 EGF-LIKE 1.
 CC DOMAIN 61 112 EGF-LIKE 2.
 CC DOMAIN 115 152 EGF-LIKE 3.
 CC DOMAIN 153 189 EGF-LIKE 4.
 CC DOMAIN 191 229 EGF-LIKE 5.
 CC DOMAIN 231 271 EGF-LIKE 6.
 CC DOMAIN 273 309 EGF-LIKE 7.
 CC DOMAIN 311 350 EGF-LIKE 8.
 CC DOMAIN 352 388 EGF-LIKE 9.
 CC DOMAIN 389 427 EGF-LIKE 10.
 CC DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 623 656 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 658 686 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 688 724 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 726 762 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 764 800 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 803 839 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 841 877 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 878 924 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 926 962 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 964 1000 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1002 1040 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1042 1081 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1083 1122 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1126 1167 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
 FT REPEAT 1168 1208 LIN/NOTCH 1.
 FT REPEAT 1209 1242 LIN/NOTCH 2.
 FT REPEAT 1243 1282 LIN/NOTCH 3.
 FT REPEAT 1628 1657 ANK 1.
 FT REPEAT 1661 1691 ANK 2.
 FT REPEAT 1695 1724 ANK 3.
 FT REPEAT 1728 1757 ANK 4.
 Query Match 1.6%; Score 8; DB 1; Length 1964;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 390 GTTPMLA 397
 Db 1696 GTTPMLA 1703
 RESULT 20
 NTCA HUMAN STANDARD; PRT; 2003 AA.
 AC 099456; 000306; Q99458; Q9H358; Q9U119; Q9U10;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (Notch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; Pubmed=9168133;
 RA Sugaya K., Sasamura S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (C/GB) polymorphism in this
 RT human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow; and Heart;
 RX MEDLINE=98360091; Pubmed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 RT sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]

RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leitman J., Ward D., Ish-Horowitz D., Attavanis-Tsakonas S.,
 RT "Human ligands of the Notch receptor."
 RL4 Am. J. Pathol. 154:785-794 (1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q09466-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q09466-2; Sequence=VSP_001406;
 CC Name=3;
 CC IsoId=Q09466-3; Sequence=VSP_001407;
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: Contains 28 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.jsb-sib.ch/announce/>
 CC or send an email to license@jsb-sib.ch).
 CC
 CC EMBL; D63395; BAA09708.1; ALT_FRAME.
 CC EMBL; D86566; BAA13116.1; -.
 CC EMBL; U95299; AAC3288.1; -.
 CC EMBL; U99335; AAC63097.1; -.
 CC EMBL; AB023861; BAB20317.1; -.
 CC EMBL; AB024520; BAA88951.1; -.
 CC EMBL; AB024578; BAA88952.1; -.
 CC HSSP; P08709; IBF9.
 CC Genew; HGNC:7884; NOTCH4.

DR MIM; 164951; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR002049; Lamthin_EGF.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PSS0088; ANK_REPEAT; 5.
 DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PSS0010; ASX_HYDROXYL; 1.
 DR PROSITE; PSS0022; EGF_1; 28.
 DR PROSITE; PSS0186; EGF_2; 21.
 DR PROSITE; PSS0187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT STGNAL 1 23
 FT CHAIN 24 2003
 FT 1432 2003
 FT
 FT CHAIN 1467 2003
 FT
 FT DOMAIN 24 1447
 FT TRANSSEM 1448 1468
 FT DOMAIN 1469 2003
 FT 24 63
 FT DOMAIN 64 115
 FT 118 155
 FT DOMAIN 156 192
 FT 194 232
 FT DOMAIN 234 274
 FT 276 312
 FT DOMAIN 314 353
 FT 355 391
 FT DOMAIN 392 430
 FT 432 473
 FT DOMAIN 475 511
 FT 513 549
 FT DOMAIN 551 587
 FT 589 625
 FT DOMAIN 626 659
 FT 661 689
 FT DOMAIN 691 727
 FT 729 765
 FT DOMAIN 767 803
 FT 806 842
 FT DOMAIN 844 880
 FT 882 928
 FT DOMAIN 930 966
 FT 968 1004
 FT DOMAIN 1006 1044
 FT 1046 1085
 FT DOMAIN 1087 1126
 FT 1130 1171
 FT DOMAIN 1172 1212
 FT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1366
 FT REPEAT 1367 1406
 FT REPEAT 1407 1447
 FT REPEAT 1448 1488
 FT REPEAT 1489 1529
 FT REPEAT 1530 1570
 FT REPEAT 1571 1611
 FT REPEAT 1612 1652
 FT REPEAT 1653 1693
 FT REPEAT 1694 1734
 FT REPEAT 1735 1775
 FT REPEAT 1776 1816
 FT REPEAT 1817 1857
 FT REPEAT 1858 1898
 FT REPEAT 1899 1939
 FT REPEAT 1940 1980
 FT REPEAT 1981 2021
 FT REPEAT 2022 2062
 FT REPEAT 2063 2103
 FT REPEAT 2104 2144
 FT REPEAT 2145 2185
 FT REPEAT 2186 2226
 FT REPEAT 2227 2267
 FT REPEAT 2268 2308
 FT REPEAT 2309 2349
 FT REPEAT 2350 2390
 FT REPEAT 2391 2431
 FT REPEAT 2432 2472
 FT REPEAT 2473 2513
 FT REPEAT 2514 2554
 FT REPEAT 2555 2595
 FT REPEAT 2596 2636
 FT REPEAT 2637 2677
 FT REPEAT 2678 2718
 FT REPEAT 2719 2759
 FT REPEAT 2760 2800
 FT REPEAT 2801 2841
 FT REPEAT 2842 2882
 FT REPEAT 2883 2923
 FT REPEAT 2924 2964
 FT REPEAT 2965 3005
 FT REPEAT 3006 3046
 FT REPEAT 3047 3087
 FT REPEAT 3088 3128
 FT REPEAT 3129 3169
 FT REPEAT 3170 3210
 FT REPEAT 3211 3251
 FT REPEAT 3252 3292
 FT REPEAT 3293 3333
 FT REPEAT 3334 3374
 FT REPEAT 3375 3415
 FT REPEAT 3416 3456
 FT REPEAT 3457 3497
 FT REPEAT 3498 3538
 FT REPEAT 3539 3579
 FT REPEAT 3580 3620
 FT REPEAT 3621 3661
 FT REPEAT 3662 3702
 FT REPEAT 3703 3743
 FT REPEAT 3744 3784
 FT REPEAT 3785 3825
 FT REPEAT 3826 3866
 FT REPEAT 3867 3907
 FT REPEAT 3908 3948
 FT REPEAT 3949 3989
 FT REPEAT 3990 4030
 FT REPEAT 4031 4071
 FT REPEAT 4072 4112
 FT REPEAT 4113 4153
 FT REPEAT 4154 4194
 FT REPEAT 4195 4235
 FT REPEAT 4236 4276
 FT REPEAT 4277 4317
 FT REPEAT 4318 4358
 FT REPEAT 4359 4399
 FT REPEAT 4400 4440
 FT REPEAT 4441 4481
 FT REPEAT 4482 4522
 FT REPEAT 4523 4563
 FT REPEAT 4564 4604
 FT REPEAT 4605 4645
 FT REPEAT 4646 4686
 FT REPEAT 4687 4727
 FT REPEAT 4728 4768
 FT REPEAT 4769 4809
 FT REPEAT 4810 4850
 FT REPEAT 4851 4891
 FT REPEAT 4892 4932
 FT REPEAT 4933 4973
 FT REPEAT 4974 5014
 FT REPEAT 5015 5055
 FT REPEAT 5056 5096
 FT REPEAT 5097 5137
 FT REPEAT 5138 5178
 FT REPEAT 5179 5219
 FT REPEAT 5220 5260
 FT REPEAT 5261 5301
 FT REPEAT 5302 5342
 FT REPEAT 5343 5383
 FT REPEAT 5384 5424
 FT REPEAT 5425 5465
 FT REPEAT 5466 5506
 FT REPEAT 5507 5547
 FT REPEAT 5548 5588
 FT REPEAT 5589 5629
 FT REPEAT 5630 5670
 FT REPEAT 5671 5711
 FT REPEAT 5712 5752
 FT REPEAT 5753 5793
 FT REPEAT 5794 5834
 FT REPEAT 5835 5875
 FT REPEAT 5876 5916
 FT REPEAT 5917 5957
 FT REPEAT 5958 5998
 FT REPEAT 6000 6040
 FT REPEAT 6041 6081
 FT REPEAT 6082 6122
 FT REPEAT 6123 6163
 FT REPEAT 6164 6204
 FT REPEAT 6205 6245
 FT REPEAT 6246 6286
 FT REPEAT 6287 6327
 FT REPEAT 6328 6368
 FT REPEAT 6369 6409
 FT REPEAT 6410 6450
 FT REPEAT 6451 6491
 FT REPEAT 6492 6532
 FT REPEAT 6533 6573
 FT REPEAT 6574 6614
 FT REPEAT 6615 6655
 FT REPEAT 6656 6696
 FT REPEAT 6697 6737
 FT REPEAT 6738 6778
 FT REPEAT 6779 6819
 FT REPEAT 6820 6860
 FT REPEAT 6861 6901
 FT REPEAT 6902 6942
 FT REPEAT 6943 6983
 FT REPEAT 6984 7024
 FT REPEAT 7025 7065
 FT REPEAT 7066 7106
 FT REPEAT 7107 7147
 FT REPEAT 7148 7188
 FT REPEAT 7189 7229
 FT REPEAT 7230 7270
 FT REPEAT 7271 7311
 FT REPEAT 7312 7352
 FT REPEAT 7353 7393
 FT REPEAT 7394 7434
 FT REPEAT 7435 7475
 FT REPEAT 7476 7516
 FT REPEAT 7517 7557
 FT REPEAT 7558 7598
 FT REPEAT 7599 7639
 FT REPEAT 7640 7680
 FT REPEAT 7681 7721
 FT REPEAT 7722 7762
 FT REPEAT 7763 7803
 FT REPEAT 7804 7844
 FT REPEAT 7845 7885
 FT REPEAT 7886 7926
 FT REPEAT 7927 7967
 FT REPEAT 7968 8008
 FT REPEAT 8009 8049
 FT REPEAT 8050 8090
 FT REPEAT 8091 8131
 FT REPEAT 8132 8172
 FT REPEAT 8173 8213
 FT REPEAT 8214 8254
 FT REPEAT 8255 8295
 FT REPEAT 8296 8336
 FT REPEAT 8337 8377
 FT REPEAT 8378 8418
 FT REPEAT 8419 8459
 FT REPEAT 8460 8500
 FT REPEAT 8501 8541
 FT REPEAT 8542 8582
 FT REPEAT 8583 8623
 FT REPEAT 8624 8664
 FT REPEAT 8665 8705
 FT REPEAT 8706 8746
 FT REPEAT 8747 8787
 FT REPEAT 8788 8828
 FT REPEAT 8829 8869
 FT REPEAT 8870 8910
 FT REPEAT 8911 8951
 FT REPEAT 8952 8992
 FT REPEAT 8993 9033
 FT REPEAT 9034 9074
 FT REPEAT 9075 9115
 FT REPEAT 9116 9156
 FT REPEAT 9157 9197
 FT REPEAT 9198 9238
 FT REPEAT 9239 9279
 FT REPEAT 9280 9320
 FT REPEAT 9321 9361
 FT REPEAT 9362 9402
 FT REPEAT 9403 9443
 FT REPEAT 9444 9484
 FT REPEAT 9485 9525
 FT REPEAT 9526 9566
 FT REPEAT 9567 9607
 FT REPEAT 9608 9648
 FT REPEAT 9649 9689
 FT REPEAT 9690 9730
 FT REPEAT 9731 9771
 FT REPEAT 9772 9812
 FT REPEAT 9813 9853
 FT REPEAT 9854 9894
 FT REPEAT 9895 9935
 FT REPEAT 9936 9976
 FT REPEAT 9977 10017
 FT REPEAT 10018 10058
 FT REPEAT 10059 10099
 FT REPEAT 10100 10140
 FT REPEAT 10141 10181
 FT REPEAT 10182 10222
 FT REPEAT 10223 10263
 FT REPEAT 10264 10303
 FT REPEAT 10305 10345
 FT REPEAT 10346 10386
 FT REPEAT 10387 10427
 FT REPEAT 10428 10468
 FT REPEAT 10469 10509
 FT REPEAT 10510 10550
 FT REPEAT 10551 10590
 FT REPEAT 10592 10632
 FT REPEAT 10633 10673
 FT REPEAT 10674 10714
 FT REPEAT 10715 10755
 FT REPEAT 10756 10796
 FT REPEAT 10797 10836
 FT REPEAT 10838 10878
 FT REPEAT 10879 10919
 FT REPEAT 10920 10960
 FT REPEAT 10961 11001
 FT REPEAT 11002 11042
 FT REPEAT 11043 11082
 FT REPEAT 11084 11124
 FT REPEAT 11125 11165
 FT REPEAT 11166 11205
 FT REPEAT 11207 11247
 FT REPEAT 11248 11288
 FT REPEAT 11289 11329
 FT REPEAT 11330 11370
 FT REPEAT 11371 11411
 FT REPEAT 11412 11452
 FT REPEAT 11453 11492
 FT REPEAT 11494 11534
 FT REPEAT 11535 11575
 FT REPEAT 11576 11615
 FT REPEAT 11617 11657
 FT REPEAT 11658 11697
 FT REPEAT 11699 11739
 FT REPEAT 11740 11780
 FT REPEAT 11781 11821
 FT REPEAT 11822 11862
 FT REPEAT 11863 11902
 FT REPEAT 11904 11944
 FT REPEAT 11945 11985
 FT REPEAT 11986 12026
 FT REPEAT 12027 12067
 FT REPEAT 12068 12108
 FT REPEAT 12109 12149
 FT REPEAT 12150 12190
 FT REPEAT 12191 12231
 FT REPEAT 12232 12272
 FT REPEAT 12273 12313
 FT REPEAT 12314 12354
 FT REPEAT 12355 12395
 FT REPEAT 12396 12436
 FT REPEAT 12437 12477
 FT REPEAT 12478 12518
 FT REPEAT 12519 12559
 FT REPEAT 12560 12600
 FT REPEAT 12601 12641
 FT REPEAT 12642 12682
 FT REPEAT 12683 12722
 FT REPEAT 12724 12764
 FT REPEAT 12765 12805
 FT REPEAT 12806 12846
 FT REPEAT 12847 12887
 FT REPEAT 12888 12928
 FT REPEAT 12929 12969
 FT REPEAT 12970 13010
 FT REPEAT 13011 13051
 FT REPEAT 13052 13092
 FT REPEAT 13093 13133
 FT REPEAT 13134 13174
 FT REPEAT 13175 13215
 FT REPEAT 13216 13256
 FT REPEAT 13257 13297
 FT REPEAT 13298 13338
 FT REPEAT 13339 13379
 FT REPEAT 13380 13420
 FT REPEAT 13421 13461
 FT REPEAT 13462 13502
 FT REPEAT 13503 13543
 FT REPEAT 13544 13584
 FT REPEAT 13585 13625
 FT REPEAT 13626 13666
 FT REPEAT 13667 13706
 FT REPEAT 13708 13748
 FT REPEAT 13749 13789
 FT REPEAT 13790 13830
 FT REPEAT 13831 13871
 FT REPEAT 13872 13912
 FT REPEAT 13913 13953
 FT REPEAT 13954 13994
 FT REPEAT 13995 14035
 FT REPEAT 14036 14076
 FT REPEAT 14077 14117
 FT REPEAT 14118 14158
 FT REPEAT 14159 14199
 FT REPEAT 14200 14240
 FT REPEAT 14241 14281
 FT REPEAT 14282 14322
 FT REPEAT 14323 14363
 FT REPEAT 14364 14404
 FT REPEAT 14405 14445
 FT REPEAT 14446 14486
 FT REPEAT 14487 14527
 FT REPEAT 14528 14568
 FT REPEAT 14569 14608
 FT REPEAT 14610 14650
 FT REPEAT 14651 14691
 FT REPEAT 14692 14732
 FT REPEAT 14733 14773
 FT REPEAT 14774 14814
 FT REPEAT 14815 14855
 FT REPEAT 14856 14896
 FT REPEAT 14897 14937
 FT REPEAT 14938 14978
 FT REPEAT 14979 15019
 FT REPEAT 15020 15060
 FT REPEAT 15061 15101
 FT REPEAT 15102 15142
 FT REPEAT 15143 15182
 FT REPEAT 15184 15224
 FT REPEAT 15225 15265
 FT REPEAT 15266 15306
 FT REPEAT 15307 15347
 FT REPEAT 15348 15388
 FT REPEAT 15389 15429
 FT REPEAT 15430 15470
 FT REPEAT 15471 15511
 FT REPEAT 15512 15552
 FT REPEAT 15553 15593
 FT REPEAT 15594 15634
 FT REPEAT 15635 15675
 FT REPEAT 15676 15715
 FT REPEAT 15717 15757
 FT REPEAT 15758 15798
 FT REPEAT 15799 15839
 FT REPEAT 15840 15880
 FT REPEAT 15881 15921
 FT REPEAT 15922 15962
 FT REPEAT 15963 16003
 FT REPEAT 16004 16044
 FT REPEAT 16045 16084
 FT REPEAT 16086 16126
 FT REPEAT 16127 16167
 FT REPEAT 16168 16208
 FT REPEAT 16209 16249
 FT REPEAT 16250 16290
 FT REPEAT 16291 16331
 FT REPEAT 16332 16372
 FT REPEAT 16373 16413
 FT REPEAT 16414 16454
 FT REPEAT 16455 16495
 FT REPEAT 16496 16536
 FT REPEAT 16537 16577
 FT REPEAT 16578 16618
 FT REPEAT 16619 16659
 FT REPEAT 16660 16700
 FT REPEAT 16701 16741
 FT REPEAT 16742 16782
 FT REPEAT 16783 16823
 FT REPEAT 16824 16864
 FT REPEAT 16865 16905
 FT REPEAT 16906 16946
 FT REPEAT 16947 16987
 FT REPEAT 16988 17028
 FT REPEAT 17029 17069
 FT REPEAT 17070 17110
 FT REPEAT 17111 17151
 FT REPEAT 17152 17192
 FT REPEAT 17193 17233
 FT REPEAT 17234 17274
 FT REPEAT 17275 17315
 FT REPEAT 17316 17356
 FT REPEAT 17357 17397
 FT REPEAT 17398 17438
 FT REPEAT 17439 17479
 FT REPEAT 17480 17520
 FT REPEAT 17521 17561
 FT REPEAT 17562 17602
 FT REPEAT 17603 17643
 FT REPEAT 17644 17684
 FT REPEAT 17685 17725
 FT REPEAT 17726 17766
 FT REPEAT 17767 17807
 FT REPEAT 17808 17848
 FT REPEAT 17849 17889
 FT REPEAT 17890 17930
 FT REPEAT 17931 17971
 FT REPEAT 17972 18012
 FT REPEAT 18013 18053
 FT REPEAT 18054 18094
 FT REPEAT 18095 18134
 FT REPEAT 18136 18176
 FT REPEAT 18177 18217
 FT REPEAT 18218 18258
 FT REPEAT 18259 18298
 FT REPEAT 18300 18340
 FT REPEAT 18341 18381
 FT REPEAT 18382 18422
 FT REPEAT 18423 18463
 FT REPEAT 18464 18504
 FT REPEAT 18505 18545
 FT REPEAT 18546 18586
 FT REPEAT 18587 18626
 FT REPEAT 18628 18668
 FT REPEAT 18669 18709
 FT REPEAT 18710 18750
 FT REPEAT 18751 18791
 FT REPEAT 18792 18832
 FT REPEAT 18833 18873
 FT REPEAT 18874 18914
 FT REPEAT 18915 18955
 FT REPEAT 18956 18996
 FT REPEAT 18997 19036
 FT REPEAT 19038 19078
 FT REPEAT 19079 19119
 FT REPEAT 19120 19160
 FT REPEAT 19161 19201
 FT REPEAT 19202 19242
 FT REPEAT 19243 19282
 FT REPEAT 19284 19324
 FT REPEAT 19325 19365
 FT REPEAT 19366 19406
 FT REPEAT 19407 19447
 FT REPEAT 19448 19488
 FT REPEAT 19489 19529
 FT REPEAT 19530 19570
 FT REPEAT 19571 19611
 FT REPEAT 19612 19652
 FT REPEAT 19653 19693
 FT REPEAT 19694 19734
 FT REPEAT 19735 19775
 FT REPEAT 19776 19816
 FT REPEAT 19817 19857
 FT REPEAT 19858 19898
 FT REPEAT 19899 19939
 FT REPEAT 19940 19980
 FT REPEAT 19981 20021
 FT REPEAT 20022 20062
 FT REPEAT 20063 20103
 FT REPEAT 20104 20144
 FT REPEAT 20145 20185
 FT REPEAT 20186 20226
 FT REPEAT 20227 20267
 FT REPEAT 20268 20308
 FT REPEAT 20309 20349
 FT REPEAT 20350 20390
 FT REPEAT 20391 20431
 FT REPEAT 20432 20472
 FT REPEAT 20473 20513
 FT REPEAT 20514 20554
 FT REPEAT 20555 20595
 FT REPEAT 20596 20636
 FT REPEAT 20637 20677
 FT REPEAT 20678 20717
 FT REPEAT 20719 20759
 FT REPEAT 20760 20800
 FT REPEAT 20801 20841
 FT REPEAT 20842 20882
 FT REPEAT 20883 20923
 FT REPEAT 20924 20964
 FT REPEAT 20965 21005
 FT REPEAT 21006 21046
 FT REPEAT 21047 21087
 FT REPEAT 21088 21128
 FT REPEAT 21129 21169
 FT REPEAT 21170 21210
 FT REPEAT 21211 21251
 FT REPEAT 21252 21292
 FT REPEAT 21293 21333
 FT REPEAT 21334 21374
 FT REPEAT 21375 21415
 FT REPEAT 21416 21456
 FT REPEAT 21457 21496
 FT REPEAT 21498 21538
 FT REPEAT 21539 21579
 FT REPEAT 21580 21620
 FT REPEAT 21621 21661
 FT REPEAT 21662 21702
 FT REPEAT 21703 21743
 FT REPEAT 21744 21784
 FT REPEAT 21785 21825
 FT REPEAT 21826 21866
 FT REPEAT 21867 21907
 FT REPEAT 21908 21948
 FT REPEAT 21949 21989
 FT REPEAT 21990 22029
 FT REPEAT 22031 22071
 FT REPEAT 22072 22112
 FT REPEAT 22113 22153
 FT REPEAT 22154 22194
 FT REPEAT 22195 22235
 FT REPEAT 22236 22276
 FT REPEAT 22277 22317
 FT REPEAT 22318 22358
 FT REPEAT 22359 22399
 FT REPEAT 22400 22440
 FT REPEAT 22441 22481
 FT REPEAT 22482 22522
 FT REPEAT 22523 22563
 FT REPEAT 22564 22604
 FT REPEAT 22605 22645
 FT REPEAT 22646 22686
 FT REPEAT 22687 22726
 FT REPEAT 22728 22768
 FT REPEAT 22769 22809
 FT REPEAT 22810 22850
 FT REPEAT

```

FT REPEAT 1733 1765 ANK 4.
FT REPEAT 1766 1798 ANK 5.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 80 BY SIMILARITY.
FT DISULFID 74 103 BY SIMILARITY.
FT DISULFID 105 114 BY SIMILARITY.
FT DISULFID 122 133 BY SIMILARITY.
FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.
FT DISULFID 205 220 BY SIMILARITY.
FT DISULFID 222 231 BY SIMILARITY.
FT DISULFID 238 249 BY SIMILARITY.

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 2003;
Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 GTTPMLA 397
Db 1701 GTTPMLA 1708

RESULT 21
PTEG_CVACA STANDARD; PRT; 35 AA.
ID PTEG_CVACA
AC OptID3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome b6-f complex subunit V (Cytochrome b6f complex subunit
DE ptegf).
GN PTEG.
OS Cyanidium caldarium.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RK-1;
RX MEDLINE=20496959; PubMed=11040290;
RA Gloechner G., Rosenthal A., Valentin K.-U.;
RT "The structure and gene repertoire of an ancient red algal plastid
RT genome."
RL J. Mol. Evol. 51:382-390(2000).
CC -1- FUNCTION: The cytochrome b6-f complex functions in the linear
CC cross-membrane transport of electrons between photosystem II and
CC I, as well as in cyclic electron flow around photosystem I. Ptegf
CC is required for either the stability or assembly of the cytochrome
CC b6-f complex (By similarity).
CC -1- SUBCELLULAR LOCATION: Thylakoid membrane-associated.
CC -1- SIMILARITY: Belongs to the ptegf family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: AF022186; AAF12884.1; -
CC HAMAP: MF_00432; -; 1.
CC InterPro: IPR003683; Cytochrome b6/f_5.
CC Pfam: PF02529; Ptegf, 1.
CC Electon transport; Chloroplast; Respiratory chain; Thylakoid;
CC Transmembrane.

```

```

FT DOMAIN 1 4 LUMINAL (POTENTIAL).
FT TRANSMEM 5 25 POTENTIAL.
FT DOMAIN 26 35 STROMAL (POTENTIAL).
SQ SEQUENCE 35 AA; 3803 MW; B03C27094A1B74F3 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 35;
Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 VLTGIV 371
Db 4 VLTGIV 10

RESULT 22
TX31_DENAN STANDARD; PRT; 80 AA.
ID TX31_DENAN
AC P18329;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Toxin C13S1C1 precursor.
OS Dendroaspis angusticeps (Eastern green mamba).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosaurs; Squamata; Scleroglossae; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Dendroaspis.
OX NCBI_TaxID=8618;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Duncanel F.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 22-80.
RC TISSUE=Venom;
RX MEDLINE=80242820; PubMed=7397227;
RA Joubert F.J., Taljaard N.;
RT "The complete primary structures of two reduced and
RT s-carboxymethylated angusticeps-type toxins from Dendroaspis
RT angusticeps (green mamba) venom."
RL Biochim. Biophys. Acta 623:449-456(1980).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) IS 40 +- 3 MG/KG BY INTRAVENOUS INJECTION.
CC -1- SIMILARITY: Belongs to the snake toxin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: X52260; CA36503.1; -
CC EMBL: X53410; CA37486.1; -
CC PIR: S10464; S10464.
CC HSBP; P01416; INTX.
CC InterPro: IPR003571; Snake_toxin.
CC Pfam: PF00087; toxin, 1.
CC ProDom: PD000206; Snake_toxin, 1.
CC PROSITE: PS00272; SNAKE_TOXIN, 1.
KW Toxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 80 TOXIN C13S1C1.
FT DISULFID 24 43 BY SIMILARITY.
FT DISULFID 38 60 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 74 79 BY SIMILARITY.
FT CONFLICT 56 56 MISSING (IN REF. 2).
SQ SEQUENCE 80 AA; 8954 MW; 5F4450DF19E721C8 CRC64;

Query Match
1.4%; Score 7; DB 1; Length 80;

```

Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLLV 294
DB 4 LLLTLLV 10

RESULT 23

TXF8_DENAN STANDARD; PRT; 81 AA.

AC P01404;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Toxin F-vIII precursor (Toxin TA2) (Toxin DBF8).
OS Dendroaspis angusticeps (Eastern green mamba).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Dendroaspis.
OX NCBI_TaxID=8618;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Ducancel F.;
RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 22-81.
RC TISSUE=Venom;
RX MEDLINE=74070661; PubMed=4203470;
RA Viljoen C.C., Botes D.P.;
RT "Snake venom toxins. The purification and amino acid sequence of
RT toxin TA2 from Dendroaspis angusticeps venom.";
RL J. Biol. Chem. 249:366-372(1974).
CC -1- FUNCTION: Its pharmacological action is unknown. It probably acts
CC synergistically with other venom components as whole venom is very
CC toxic.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) is >250 mg/kg by subcutaneous injection.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X53409; CAA37485.1; -.
DR PIR; S21289; T5EP2A.
DR HSSP; P01416; INTX.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Toxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 81
FT DISULFID 24 43
FT DISULFID 38 60
FT DISULFID 62 73
FT DISULFID 74 79
FT CONFLICT 54 54
SQ SEQUENCE 81 AA; 8849 MW; 733195E17A656C97 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLLV 294
DB 4 LLLTLLV 10

DB 4 LLLTLLV 10

RESULT 24

CBTB_NAJAT STANDARD; PRT; 82 AA.

AC P80958; O42285;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cobrotoxin b precursor (CBT b) (Short neurotoxin).
OS Naja atra (Chinese cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8656;
RN [1]
RP SEQUENCE FROM N.A., AND REVISION TO 67.
RC TISSUE=Liver, and Venom;
RX MEDLINE=98158338; PubMed=9496573;
RA Chang L.-S., Chou Y.-C., Lin S.-R., Wu B.-N., Lin J., Hong E.,
RA Sun Y.-J., Hsiao C.-D.;
RT "A novel neurotoxin, cobrotoxin b, from Naja naja atra (Taiwan cobra)
RT venom: purification, characterization, and gene organization.";
RL J. Biochem. 122:1252-1259(1997).

CC -1- FUNCTION: Produces peripheral paralysis by blocking neuromuscular
CC transmission at the postsynaptic site. Binds to the nicotinic
CC acetylcholine receptor (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Y13399; CAA73829.2; -.
DR EMBL; AF031472; AAB86636.1; -.
DR PIR; JCS892; JCS892.
DR HSSP; P01427; INOR.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Toxin; Neurotoxin; Postsynaptic neurotoxin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 82
FT DISULFID 24 44
FT DISULFID 38 61
FT DISULFID 63 74
FT DISULFID 75 80
SQ SEQUENCE 82 AA; 9139 MW; 1FPA21189C08BEB8 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLLTLLV 294
DB 4 LLLTLLV 10

RESULT 25
CBTA_NAJAT

ID CHTA NAJAT STANDARD; PRT; 83 AA.
AC P01430.013079; Created
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cobrotoxin precursor (CBT) (Short neurotoxin 1).
OS Naja atra (Chinese cobra), and
OS Naja kaouthia (Monocled cobra) (Naja siamensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8656, 8649;
[1]
RE SEQUENCE FROM N.A.
RE SPECIES=N. atra; TISSUE=Venom gland;
RA Chu R.C.; Yang C.C.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
[2]
RE SEQUENCE FROM N.A.
RE SPECIES=N. atra; TISSUE=Liver;
RX MEDLINE=9804246; PubMed=9367842;
RA Chang L.-S., Lin J., Chou Y.-C., Hong E.;
RT "Genomic structures of cardiotoxin 4 and cobrotoxin from Naja naja atra (Taiwan cobra).";
RL Biochem. Biophys. Res. Commun. 239:756-762(1997).
[3]
RE SEQUENCE OF 22-83.
RE SPECIES=N. atra; TISSUE=Venom;
RX MEDLINE=69294410; PubMed=5820687;
RA Yang C.C., Yang H.J., Huang J.S.;
RT "The amino acid sequence of cobrotoxin";
RL Biochim. Biophys. Acta 188:65-77(1999).
[4]
RE SEQUENCE OF 22-83.
RE SPECIES=N. atra; TISSUE=Venom;
RX MEDLINE=93192268; PubMed=8448165;
RA Chlou S.H., Raynor R.L., Zheng B., Chambers T.C., Kuo J.F.;
RT "Cobra venom cardiotoxin (cytotoxin) isoforms and neurotoxin: comparative potency of protein kinase C inhibition and cancer cell cytotoxicity and modes of enzyme inhibition.";
RL Biochemistry 32:2062-2067(1993).
[5]
RE SEQUENCE OF 22-83.
RE SPECIES=N. kaouthia; TISSUE=Venom;
RX MEDLINE=22035208; PubMed=12039691;
RA Meng Q.-X., Wang W.-Y., Lu Q.-M., Jin Y., Wei J.-F., Zhu S.-W., Xiong Y.-L.;
RT "A novel short neurotoxin, cobrotoxin C, from monocellate cobra (Naja kaouthia) venom: isolation and purification, primary and secondary structure determination, and tertiary structure modeling.";
RL Comp. Biochem. Physiol. 132C:113-121(2002).
[6]
RE DISULFIDE BONDS.
RE SPECIES=N. atra;
RX MEDLINE=71125117; PubMed=5533659;
RA Yang C.C., Yang H.J., Chiu R.H.C.;
RT "The position of disulfide bonds in cobrotoxin.";
RL Biochim. Biophys. Acta 214:355-363(1970).
[7]
RE STRUCTURE BY NMR.
RE SPECIES=N. atra;
RX MEDLINE=91065386; PubMed=2249693;
RA Yu C., Lee C.-S., Chuang L.-C., Shet Y.-R., Wang C.Y.;
RT "Two-dimensional NMR studies and secondary structure of cobrotoxin in aqueous solution.";
RL Eur. J. Biochem. 193:789-799(1990).
[8]
RE STRUCTURE BY NMR.
RE SPECIES=N. atra;
RX MEDLINE=93183875; PubMed=8443154;
RA Yu C., Bhaskaran R., Chuang L.-C., Yang C.C.;
RT "Solution conformation of cobrotoxin: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study.";

RL Biochemistry 32:2131-2136(1993).
CC -1- FUNCTION: Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. Binds to the nicotinic acetylcholine receptor.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- MISCELLANEOUS: LD(50) is 0.09 mg/kg by subcutaneous injection.
CC -1- SIMILARITY: Belongs to the snake toxin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U58519; AAB03221.1; -;
CC EMBL; U58520; AAB03222.1; -;
CC EMBL; U58521; AAB03223.1; -;
CC EMBL; U77490; AAB36830.1; -;
CC EMBL; U77491; AAB36931.1; -;
CC EMBL; U77492; AAB36932.1; -;
CC EMBL; U42582; AAB01538.1; -;
CC EMBL; Y12492; CAA73097.2; -;
CC PIR; JCS769; NINJIF.
CC PDB; 1COD; 20-JUL-95.
CC PDB; 1COE; 20-JUL-95.
CC InterPro; IPR003571; Snake_toxin.
CC Pfam; PF00087; toxin; 1.
CC Prodom; PD000206; Snake_toxin; 1.
CC PROSITE; PS00272; SNAKE_TOXIN; 1.
KM Toxin; Neurotoxin; Postsynaptic neurotoxin; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 83 COBROTOXIN.
FT DISULFID 24 45
FT DISULFID 38 62
FT DISULFID 64 75
FT DISULFID 76 81
FT STRAND 24 27
FT STRAND 33 36
FT STRAND 45 50
FT TURN 53 54
FT STRAND 57 59
FT STRAND 62 62
FT STRAND 72 76
SQ SEQUENCE 83 AA; 9261 MW; 4DD6077C92717052 CRC64;
Query March 1.4%; Score 7; DB 1; Length 83;
Best local similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 288 LLTLTLV 294
DB 4 LLTLTLV 10
RESULT 26
HG15 CHICK STANDARD; PRT; 104 AA.
AC P12902;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nonhistone chromosomal protein HMG-14A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
[1]
RE SEQUENCE FROM N.A.


```

RX MEDLINE=88255874; PubMed=3384337;
RA Dodgson J.B., Browne D.L., Black A.J.;
RT "Chicken chromosomal protein HMG-14 and HMG-17 cDNA clones:
RT isolation, characterization and sequence comparison.";
RL Gene 63:287-295(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE=93185924; PubMed=8444343;
RA Browne D.L., Dodgson J.B.;
RT "The gene encoding chicken chromosomal protein HMG-14a is transcribed
RT into multiple mRNAs.";
RL Gene 124:199-206(1993).
CC -1- FUNCTION: BIND TO THE INNER SIDE OF THE NUCLEOSOMAL DNA THUS
CC ALTERING THE INTERACTION BETWEEN THE DNA AND THE HISTONE OCTAMER.
CC MAY BE INVOLVED IN THE PROCESS WHICH MAINTAINS TRANSCRIBABLE GENES
CC IN A UNIQUE CHROMATIN CONFORMATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: THERE ARE TWO HMG-14 PROTEINS IN CHICKEN: HMG-14A
CC -1- THE MAJOR COMPONENT, AND HMG-14B THE MINOR COMPONENT.
CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; M26675; AAA48815.1; -
DR EMBL; X63083; CAA44804.1; -
DR EMBL; X63084; CAA44804.1; JOINED.
DR EMBL; X63085; CAA44804.1; JOINED.
DR EMBL; X63086; CAA44804.1; JOINED.
DR EMBL; X63087; CAA44804.1; JOINED.
DR PIR; S22122; NSCHM4.
DR InterPro: IPR000079; HMG_14_17.
DR Pfam; PF01101; HMG14_17; 1.
DR PRINTS; PR00925; NONHISHMG17.
DR SMART; SM00527; HMG17; 1.
DR PROSITE; PS00355; HMG14_17; 1.
KW Nuclear protein; DNA-binding; Multigene family.
FT INIT MET 0
SQ SEQUENCE 104 AA; 11225 MW; SEC6F5F165611316 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 KKKKGAK 269
DB 58 KKKKGAK 64

RESULT 27
RL3E_SULSO STANDARD; PRT; 108 AA.
AC Q980R3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 508 ribosomal protein L30e.
GN RPL30E OR SSO0221.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_Taxid=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

```

```

RA Aways M.J., Chan-Wehner C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Tni-Vyoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- SIMILARITY: BELONGS TO THE L30E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; AE006659; AAK40564.1; -
DR PIR; A90163; E90163.
DR HAMAP; MF_00481; 1.
DR InterPro: IPR000231; Ribosomal_L30e.
DR PIR; IPR004038; Ribosomal_L7A.
DR Pfam; PF01248; Ribosomal_L7ae; 1.
DR ProDom; PD004495; Ribosomal_L30e; 1.
DR PROSITE; PS00709; RIBOSOMAL_L30E_1; 1.
DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 108 AA; 12095 MW; 4B24861EC334AA2C CRC64;

Query Match 1.4%; Score 7; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 YAKLSNI 88
DB 60 YAKLSNI 66

RESULT 28
RL3E_METUA STANDARD; PRT; 110 AA.
AC P54051;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 505 ribosomal protein L30e.
GN RPL30E OR MJ1044.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE L30E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U67547; AAB99048.1; -
 DR PIR; C64430; C64430.
 DR HSSP; P14120; ICN9.
 DR TIGR; M1044; -
 DR HAMAP; MF_00481; -; 1.
 DR InterPro; IPR00231; Ribosomal_L30e.
 DR InterPro; IPR004038; Ribosomal_L7A.
 DR Pfam; PF01248; Ribosomal_L7Ae; 1.
 DR ProDom; PD004495; Ribosomal_L30e; 1.
 DR PROSITE; PS00709; RIBOSOMAL_L30E_1; 1.
 DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
 KM Ribosomal protein; Complete proteome.
 SQ SEQUENCE 110 AA; 12134 MW; EBCCA4E135E098B6 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 YAKLSNI 88
 |||||
 Db 58 YAKLSNI 64

RESULT 29
 HBBO_PAGBO STANDARD; PRT; 146 AA.
 ID HBBO_PAGBO
 AC P82345;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Hemoglobin beta-0 chain (HB 0).
 GN HBBO.
 OS Pagothenia borchgrevinkii (Bald rockcod) (Trematomus borchgrevinkii).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
 OC Nototheniidae; Nototheniidae; Pagothenia.
 OX NCBI_TaxID=8213;
 RN [1]
 RP SEQUENCE.
 RC Tissue-Blood;
 RA Riccio A., Tamburrini M., Carratore V., di Prisco G.;
 RT "Functionally distinct haemoglobins of the cryopelagic antarctic teleost Pagothenia borchgrevinkii.";
 RL J. Fish Biol. 57:20-32(2000).
 CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE VARIOUS PERIPHERAL TISSUES.
 CC -1- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.
 CC -1- TISSUE SPECIFICITY: Red blood cells.
 CC -1- MISCELLANEOUS: THIS FISH HAS FIVE HEMOGLOBINS: HB C, HB O, HB 1, HB 2 AND HB 3. HB 0 PRESENTS THE STRONGEST BOHR EFFECT WHILE HB 1 PRESENTS THE WEAKEST BOHR EFFECT.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR HSSP; P80044; 1PBX.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; globin; 1.
 DR PROSITE; PS01033; GLOBIN; 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16364 MW; 1DC991617574CF73 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 30
 HBBC_PAGBE STANDARD; PRT; 146 AA.
 ID HBBC_PAGBE
 AC P45722;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin beta-C chain.
 GN HBBC.
 OS Pagothenia bernacchii (Emerald rockcod) (Trematomus bernacchii).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
 OC Nototheniidae; Nototheniidae; Trematomus.
 OX NCBI_TaxID=40690;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94193650; PubMed=8144556;
 RA D'Avino R., Caruso C., Tamburrini M., Romano M., Rutigliano B.,
 RA Polverino de Laureto P., Camardella L., Carratore V., di Prisco G.;
 RT "Molecular characterization of the functionally distinct hemoglobins of the Antarctic fish Trematomus newnesi.";
 RL J. Biol. Chem. 269:9675-9681(1994).
 CC -1- FUNCTION: COMPONENT OF HEMOGLOBIN C, A MINOR GLOBIN OF THIS FISH.
 CC -1- SUBUNIT: HEMOGLOBIN C (HB C) IS A TETRAMER OF TWO ALPHA CHAINS AND TWO BETA-C CHAINS.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR HSSP; P80044; 1PBX.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; globin; 1.
 DR PROSITE; PS01033; GLOBIN; 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16351 MW; 990E4DA62AF046 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 31
 HBBC_TRENE STANDARD; PRT; 146 AA.
 ID HBBC_TRENE
 AC P45721;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemoglobin beta-C chain.
 GN HBBC.
 OS Trematomus newnesi (Dusky notothen).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
 OC Nototheniidae; Nototheniidae; Trematomus.
 OX NCBI_TaxID=35730;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94193650; PubMed=8144556;
 RA D'Avino R., Caruso C., Tamburrini M., Romano M., Rutigliano B.,
 RA Polverino de Laureto P., Camardella L., Carratore V., di Prisco G.;
 RT "Molecular characterization of the functionally distinct hemoglobins of the Antarctic fish Trematomus newnesi.";
 RL J. Biol. Chem. 269:9675-9681(1994).
 CC -1- FUNCTION: COMPONENT OF HEMOGLOBIN C, A MINOR (20% TO 25%) GLOBIN OF THIS FISH.
 CC -1- SUBUNIT: HEMOGLOBIN C (HB C) IS A TETRAMER OF TWO ALPHA-1

CC CHAINS AND TWO BETA-C CHAINS.
 CC -1- SIMILARITY: Belongs to the globin family.
 DR PIR, D54403; D54403.
 DR HSSB, P80044; 1PBX.
 DR InterPro; IPR002337; Beta_haem.
 DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; Globin; 1.
 DR PRINTS; PR00814; BETAHAEM.
 DR PROSITE; PS01033; GLOBIN; 1.
 KM Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 146 AA; 16260 MW; 9C05E8B6E092979 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 441 VVAARFG 447
 |||||
 Db 113 VVAARFG 119

RESULT 32
 NU6M LOBL STANDARD; PRT; 168 AA.
 AC 047478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN Ndc.
 OS Loligo bleekeri (Bleeker's squid).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
 OC Decapodiformes; Loliginidae; Loligo.
 OX NCBI_TaxID=6617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tomita K., Ueda T., Watanabe K.;
 RT "Completing of squid (Loligo bleekeri) mitochondrial genome
 sequencing.";
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB009838; BAA24061.1; -
 DR EMBL; AB029616; BAB03648.1; -
 DR InterPro; IPR001457; Oxidored_q3.
 DR Pfam; PF00499; oxidored_q3; 1.
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 168 AA; 19301 MW; DBA1A1726BF5CEAB CRC64;

Query Match 1.4%; Score 7; DB 1; Length 168;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 472 LAVILFL 478
 |||||
 Db 140 LAVILFL 146

RESULT 33
 NU6M CYPCA STANDARD; PRT; 172 AA.
 AC P24982;

DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN MTD6 OR ND6.
 OS Cyprinus carpio (Common carp).
 CC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Cyprinus.
 OX NCBI_TaxID=7962;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94223691; PubMed=8169959;
 RX Chang Y.S., Huang P.L., Lo T.B.;
 RT "The complete nucleotide sequence and gene organization of carp
 RT (Cyprinus carpio) mitochondrial genome.";
 RL J. Mol. Evol. 38:138-155(1994).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X61010; CAA43338.1; -
 DR EMBL; X61010; CAA43338.1; -
 DR PIR; S36007; S36007.
 DR InterPro; IPR001457; Oxidored_q3.
 DR Pfam; PF00499; oxidored_q3; 1.
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 172 AA; 17971 MW; B3EBBADA8B88558 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 172;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 288 LLTLTLV 294
 |||||
 Db 150 LLTLTLV 156

RESULT 34
 NU6M CARAU STANDARD; PRT; 173 AA.
 AC 078689;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
 GN MTD6 OR ND6.
 OS Carassius auratus (Goldfish).
 CC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Carassius.
 OX NCBI_TaxID=7957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A23 / Langsdorff; TISSUE=Oocyte;
 RA Murakami M., Yamashita Y., Fujitani H.;
 RT "The complete sequence of mitochondrial genome from a gynogenetic
 RT triploid 'gibuna' (Carassius auratus langsdorff).";
 RL Zool. Sci. 15:335-337(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cv1er1;
 RA Murakami M.;
 RT "Carassius auratus cv1er1 mitochondrial DNA, complete sequence.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB06953; BAA31249.1; -
CC EMBL; AB045144; BAB40359.1; -
CC InterPro; IPR001457; Oxidored_q3.
CC Pfam; PF00499; oxidored_q3; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 173 AA; 18095 MW; 3F4E4B5AB50C30C CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 LLLTLV 294
Db 151 LLLTLV 157

RESULT 35
ID NUGM_GADMO STANDARD; PRT; 173 AA.
AC P55783;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
GN MTND6 OR ND6.
OS Gadus morhua (Atlantic cod).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OC NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Norwegian coastal 1;
RX MEDLINE=96414925; PubMed=8817926;
RA Johansen S., Bakke I.;
RT "The complete mitochondrial DNA sequence of Atlantic cod (Gadus
RT morhua): relevance to taxonomic studies among codfishes.";
RL Mol. Mar. Biol. Biotechnol. 5:203-214(1996).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99772; CAA68116.1; -
CC PIR; T11831; T11831.
CC InterPro; IPR001457; Oxidored_q3.
CC Pfam; PF00499; oxidored_q3; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
CC SEQUENCE 173 AA; 18090 MW; 7E5754621B4C7FCB CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 LLLTLV 294
Db 151 LLLTLV 157

```

```

RESULT 36
ID RRF_WIGBR STANDARD; PRT; 185 AA.
AC Q8D2G5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN RRF OR WIGBR3890.
OS Wigglesworthia glossinidia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
CC NCBI_TaxID=356970;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akayan L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Akao S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, Wigglesworthia glossinidia."
RL Nat. Genet. 32:402-407(2002).
CC -1- FUNCTION: Responsible for the release of ribosomes from messenger
CC RNA at the termination of protein biosynthesis. May increase the
CC efficiency of translation by recycling ribosomes from one round of
CC translation to another (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB063522; BAC24535.1; -
CC HAMAP; MF_00040; -; 1
CC InterPro; IPR002661; RRF.
CC Pfam; PF01765; RRF; 1.
CC PRODOM; PD004103; RRF; 1.
CC TRIFRAMS; TRIGR00496; frr; 1.
CC Protein biosynthesis; Complete proteome.
CC SEQUENCE 185 AA; 21650 MW; C71559E7155DECD CRC64;
QY Query Match 1.4%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 KLSNLS 90
Db 53 KLSNLS 59

RESULT 37
ID RK24_SPTOL STANDARD; PRT; 191 AA.
AC P27683;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 50S ribosomal protein L24, chloroplast precursor (Clr24).
GN RPL24.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
CC NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91365238; PubMed=1889743;
RA Carol P., Li Y.F., Maché R.;

```

```

RT "Conservation and evolution of the nucleus-encoded and chloroplast-
RT specific ribosomal proteins in pea and spinach."
RL Gene 103:139-145(1991).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: MS8522; AAA34042.1; -.
DR PIR: JH0585; JH0585.
DR InterPro: IPR005824; KOW.
DR InterPro: IPR006646; KOW_sub.
DR InterPro: IPR003256; Ribosomal_L24.
DR InterPro: IPR005825; Ribosomal_L24_26.
DR Pfam: PF00467; KOW: 1.
DR ProDom: PD001677; Ribosomal_L24; 1.
DR SMART: SM00739; KOW: 1.
DR TIGRFAMs: TIGR01079; rplX_bact; 1.
DR PROSITE: PS01108; RIBOSOMAL_L24; 1.
KM Ribosomal protein; Chloroplast; Transit peptide.
FT TRANSIT 1 45 CHLOROPLAST.
FT CHAIN 46 191 50S RIBOSOMAL PROTEIN L24.
FT SEQUENCE 191 AA; 21439 MW; 1BE01D8355824E7D CRC64;

Query March 1.4%; Score 7; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 LKEQEVA 506
Db 138 LKEQEVA 144

RESULT 38
INF_MEIGA STANDARD; PRT; 192 AA.
ID INF_MEIGA
AC P51527;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Interferon precursor.
OS Melaleuca leucadendron (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nicholas;
RX MEDLINE=96079079; PubMed=7494342;
RA Suresh M., Karaca K., Foster D., Sharma J.M.;
RT "Molecular and functional characterization of turkey interferon."
RL J. Virol. 69:8159-8163(1995).
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28140; AAB40029.1; -.
DR InterPro: IPR000471; Interferon_abd.
DR Pfam: PF00143; Interferon; 1.
DR ProDom: PD000550; Interferon_abd; 1.

```

```

DR SMART: SM00076; IFabd; 1.
DR PROSITE: PS00252; INTERFERON A B D; 1.
KM Cytochrome; Antiviral; Glycoprotein; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 192 INTERFERON.
FT DISULFID 31 128 BY SIMILARITY.
FT DISULFID 67 167 BY SIMILARITY.
FT DISULFID 60 154 POTENTIAL.
FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 192 AA; 21884 MW; 9ED81EF2DC4FF2C3 CRC64;

Query March 1.4%; Score 7; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILILLTLL 293
Db 15 ILILLTLL 21

RESULT 39
INF_CHICK STANDARD; PRT; 193 AA.
ID INF_CHICK
AC P42165;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Interferon type A1/A2 precursor.
GN IFNA1 AND IFNA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9435855; PubMed=7521382;
RA Sekellick M.J., Ferrandino A.F., Hopkins D.A., Marcus P.I.;
RT "Chicken interferon gene: cloning, expression, and analysis."
RL J. Interferon Res. 14:71-79(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Spleen;
RX MEDLINE=96205950; PubMed=8631799;
RA Sick C., Schultz U., Staehel P.;
RT "A family of genes coding for two serologically distinct chicken
RT interferons."
RL J. Biol. Chem. 271:7635-7639(1996).
CC -1- FUNCTION: HAS ANTI-VIRAL ACTIVITIES.
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- DEVELOPMENTAL STAGE: APPEARS FIRST AT 3 HOURS POST-INFECTION,
CC INCREASES TO GIVE THE STRONGEST SIGNAL AT ABOUT 9 HOURS AND
CC GRADUALLY WANES TO ALMOST NOTHING AT 24 HOURS.
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07868; AAA50213.1; -.
DR EMBL: X92477; CAA63215.1; -.
DR EMBL: X92476; CAA63214.1; -.
DR PIR: I50693; I50693.
DR InterPro: IPR000471; Interferon_abd.
DR Pfam: PF00143; Interferon; 1.
DR PRINTS: PR00266; INTERFERONAB.
DR ProDom: PD000550; Interferon_abd; 1.

```

```

DR SMART; SMO0076; IFabd; 1.
DR PROSITE; PS00252; INTERFERON A B D; 1.
KW Cycokine; Antiviral; Glycoprotein; signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 193
FT DISULFID 32 129
FT DISULFID 61 155
FT DISULFID 68 168
FT CARBOHYD 65 65
FT CARBOHYD 71 71
FT CARBOHYD 108 108
FT CARBOHYD 186 186
FT VARIANT 65 65
SQ SEQUENCE 193 AA; 22116 MM; 35AF74479C450A15 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 193;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTL 293
DB 15 ILTLTL 21

RESULT 40
INF3 CHICK STANDARD; PRT; 193 AA.
AC Q90872;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Interferon type A3 precursor.
GN IFNA3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Spleen;
RX MEDLINE=96205950; PubMed=8631799;
RA Sick C.; Schultz U.; Staeheli P.;
RT "A family of genes coding for two serologically distinct chicken
interferons." 271:7635-7639(1996).
RL J. Biol. Chem. 271:7635-7639(1996).
CC -1- FUNCTION: HAS ANTIVIRAL ACTIVITIES.
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92478; CAA63216.1; -
DR InterPro; IPR000471; Interferon_abd.
DR Pfam; PF00143; Interferon; 1.
DR PRINTS; PR00266; INTERFERONAB.
DR ProDom; PD000550; Interferon_abd; 1.
DR SMART; SMO0076; IFabd; 1.
DR PROSITE; PS00252; INTERFERON A B D; 1.
KW Cycokine; Antiviral; Glycoprotein; signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 193
FT DISULFID 32 129
FT DISULFID 61 155
FT DISULFID 68 168
FT CARBOHYD 65 65
SQ SEQUENCE 193 AA; 22116 MM; 35AF74479C450A15 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 193;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTL 293
DB 15 ILTLTL 21

Search completed: November 25, 2003, 10:05:47
Job time : 23 secs

```

```

FT CARBOHYD 71 71
FT CARBOHYD 108 108
FT CARBOHYD 186 186
SQ SEQUENCE 193 AA; 22166 MM; B53375279C530B7B CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 193;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTL 293
DB 15 ILTLTL 21

Search completed: November 25, 2003, 10:05:47
Job time : 23 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:32:01 ; Search time 64 Seconds

(without alignments)
2076.517 Million cell updates/sec

Title: US-09-869-433-2
Perfect score: 515
Sequence: 1 MTKTEKPEFKLRSFLMPFH.....AQSAKQEVAAQSDAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: SPREMBL_23:*
2: sp_archaea:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	5.6	528	2 Q9SEV3	Q9SEV3 chlamydia t
2	23	2.5	501	16 Q92HP9	Q92HP9 rickettsia
3	11	2.1	105	10 Q945F3	Q945F3 medicago sa
4	11	2.1	588	8 Q8MPR8	Q8MPR8 citrus hybr
5	10	1.9	511	16 Q92HV4	Q92HV4 rickettsia
6	10	1.9	514	2 Q8KPN2	Q8KPN2 holospira o
7	10	1.9	644	10 Q9AS91	Q9AS91 oryza sativ
8	9	1.7	463	16 P95024	P95024 mycobacteri
9	9	1.7	463	16 Q8VUG6	Q8VUG6 mycobacteri
10	9	1.7	498	16 Q92J16	Q92J16 rickettsia
11	9	1.7	671	8 Q8MHS7	Q8MHS7 galdieria s
12	8	1.6	111	17 Q9YFB6	Q9YFB6 aeropyrum p
13	8	1.6	115	4 Q8NFV3	Q8NFV3 homo sapien
14	8	1.6	122	4 Q9BV22	Q9BV22 homo sapien
15	8	1.6	154	11 Q95362	Q95362 mus musculu
16	8	1.6	173	8 Q94592	Q94592 elassoma ev

17	8	1.6	194	16 Q24960	Q24960 helicobacte
18	8	1.6	197	16 Q9ZMT1	Q9ZMT1 helicobacte
19	8	1.6	217	4 Q96R14	Q96R14 homo sapien
20	8	1.6	238	4 Q81U63	Q81U63 homo sapien
21	8	1.6	245	6 Q95J13	Q95J13 pan troglod
22	8	1.6	247	15 Q85089	Q85089 primate t-1
23	8	1.6	282	17 Q9YFUI	Q9YFUI aeropyrum p
24	8	1.6	284	11 Q8K125	Q8K125 mus musculu
25	8	1.6	306	11 Q8VP92	Q8VP92 mus musculu
26	8	1.6	310	11 Q9WV14	Q9WV14 mus musculu
27	8	1.6	310	11 Q8VFB2	Q8VFB2 mus musculu
28	8	1.6	312	11 Q9WV13	Q9WV13 mus musculu
29	8	1.6	312	11 Q9WV11	Q9WV11 mus musculu
30	8	1.6	312	11 Q9R022	Q9R022 mus musculu
31	8	1.6	340	16 Q8F697	Q8F697 leptospira
32	8	1.6	342	4 Q9P042	Q9P042 homo sapien
33	8	1.6	356	4 Q9U021	Q9U021 homo sapien
34	8	1.6	356	4 Q96FY2	Q96FY2 homo sapien
35	8	1.6	357	4 Q60376	Q60376 homo sapien
36	8	1.6	360	10 Q9S1R6	Q9S1R6 arabidopsis
37	8	1.6	400	2 Q8GLH2	Q8GLH2 photorhabd
38	8	1.6	480	16 Q8YV45	Q8YV45 anabaena sp
39	8	1.6	485	6 Q9GLN8	Q9GLN8 pan troglod
40	8	1.6	485	6 Q9GLP7	Q9GLP7 pan troglod
41	8	1.6	485	6 Q9GLP6	Q9GLP6 gorilla gor
42	8	1.6	492	15 Q8AGK3	Q8AGK3 simian t-1y
43	8	1.6	510	16 Q85091	Q85091 primate t-1
44	8	1.6	510	16 Q8F082	Q8F082 corynebacte
45	8	1.6	520	11 Q921T2	Q921T2 mus musculu
46	8	1.6	545	10 Q8GMC3	Q8GMC3 arabidopsis
47	8	1.6	921	5 Q8MR77	Q8MR77 drosophila
48	8	1.6	967	2 Q8RK80	Q8RK80 mycobacteri
49	8	1.6	1013	3 Q9HF64	Q9HF64 ashbya goss
50	8	1.6	1360	5 Q9VD44	Q9VD44 drosophila
51	8	1.6	1772	16 Q9X877	Q9X877 streptomyce
52	8	1.6	3152	12 Q8U237	Q8U237 leek yellow
53	8	1.4	49	15 Q87390	Q87390 chimpanzee
54	8	1.4	58	16 Q8DS81	Q8DS81 streptococ
55	8	1.4	62	12 Q8QL33	Q8QL33 sulfolobus
56	8	1.4	84	13 Q9W6X0	Q9W6X0 najia atra (
57	8	1.4	83	13 Q57326	Q57326 najia sputat
58	8	1.4	83	13 Q9YGJ6	Q9YGJ6 najia sputat
59	8	1.4	83	13 Q9DE57	Q9DE57 najia atra (
60	8	1.4	83	13 Q9YJG5	Q9YJG5 najia sputat
61	8	1.4	83	13 Q9PTF0	Q9PTF0 najia atra (
62	8	1.4	84	13 Q57327	Q57327 najia sputat
63	8	1.4	84	16 Q99YR7	Q99YR7 streptococ
64	8	1.4	84	16 Q8P038	Q8P038 streptococ
65	8	1.4	84	16 Q8G6M5	Q8G6M5 streptococ
66	8	1.4	92	10 Q9ZTM9	Q9ZTM9 petunia hyb
67	8	1.4	101	8 Q99713	Q99713 neomorphus
68	8	1.4	102	2 Q52201	Q52201 mycobacteri
69	8	1.4	103	2 Q9ZNM6	Q9ZNM6 onion yellow
70	8	1.4	104	2 Q8VUR9	Q8VUR9 aster yellow
71	8	1.4	106	13 Q91988	Q91988 xenopus lae
72	8	1.4	106	13 Q96UF2	Q96UF2 xenopus lae
73	8	1.4	107	13 Q9PTT1	Q9PTT1 gallus gall
74	8	1.4	114	16 Q8PC77	Q8PC77 xanthomonas
75	8	1.4	116	16 Q86364	Q86364 mycobacteri
76	8	1.4	124	2 Q8GCE1	Q8GCE1 streptomyce
77	8	1.4	126	11 Q8BS28	Q8BS28 mus musculu
78	8	1.4	128	10 Q9M5M4	Q9M5M4 euphorbia e
79	8	1.4	136	5 Q8MNV2	Q8MNV2 caenorhabdi
80	8	1.4	137	16 Q8EYV1	Q8EYV1 leptospira
81	8	1.4	148	8 Q9MD16	Q9MD16 scenedesmus
82	8	1.4	148	16 Q8EYV3	Q8EYV3 leptospira
83	8	1.4	152	17 Q58535	Q58535 pyrococcus
84	8	1.4	155	2 P72567	P72567 streptococ
85	8	1.4	155	16 Q9A1M6	Q9A1M6 streptococ
86	8	1.4	156	2 Q93QW8	Q93QW8 pseudomonas
87	8	1.4	157	12 Q83079	Q83079 lychnis rin
88	8	1.4	158	16 Q8ET47	Q8ET47 oceanoplacil
89	8	1.4	159	8 Q8S1L8	Q8S1L8 scaphidura

90 1.4 161 5. 09VT62
91 1.4 166 2. 08X05
92 7 1.4 168 16. 08G00
93 7 1.4 173 8. 08WB67
94 7 1.4 173 8. 08LUL1
95 7 1.4 173 8. 0951J1
96 7 1.4 173 8. 0951I9
97 7 1.4 173 8. 08WB69
98 7 1.4 173 8. 094SD1
99 7 1.4 173 8. 094Q0
100 7 1.4 173 8. 08H81

ALIGNMENTS

RESULT 1

Q986V3 PRELIMINARY; PRT; 528 AA.
AC Q986V3; PRELIMINARY; PRT; 528 AA.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nucleoside triphosphate transport protein 1.
GN NP1.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99138740; PubMed=9973346;
RA Tjaden J., Winkler H.H., Schweeppe C., van der Laan M., Moehlmann T.,
Neuhaus E.;
RT "Two nucleoside transport proteins in Chlamydia trachomatis : One for
net nucleoside triphosphate uptake and the other for the transport of
energy.";
RT J. Bacteriol. 181:1196-1202 (1999).
RL EMBL; AJ010586; CAB39534.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR InterPro; IPR000531; TOMB_boxC.
DR Pfam; PF03219; TMC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
SQ SEQUENCE 528 AA; 58099 MW; 8FECDD5FC08B0E3 CRC64;

Query Match 5.6%; Score 29; DB 2; Length 528;
Best Local Similarity 100.0%; Pred. No. 7.1e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AYIPLDQKXKGAIDVVAARFGSGG 451
DB 421 AYIPLDQKXKGAIDVVAARFGSGG 449

RESULT 2

Q92HP9 PRELIMINARY; PRT; 501 AA.
AC Q92HP9; PRELIMINARY; PRT; 501 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADP carrier protein.
GN TMC3 OR RC0722.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiasee; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RA MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,

RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098 (2001).
DR EMBL; AE008650; ALU03260.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TMC; 1.
DR TIGRFAMs; TIGR00769; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 501 AA; 57104 MW; 2EC589FAD865EA3B CRC64;

Query Match 2.5%; Score 13; DB 16; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 LFDSTKEMAYIPL 427
DB 415 LFDSTKEMAYIPL 427

RESULT 3

Q945F3 PRELIMINARY; PRT; 105 AA.
AC Q945F3; PRELIMINARY; PRT; 105 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADP/ATP translocase (Fragment).
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Reagen S; TISSUE=Callus;
RA Winicov I.;
RT "Fragment of putative chloroplast ADP/ATP translocase from salt
tolerant callus.";
RT Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
RL EMBL; AF16339; ALU0405.1; -
DR InterPro; IPR004667; ADP_ATP_car.
DR Pfam; PF03219; TMC; 1.
FT NON_TER 1 1
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 11212 MW; 1D0B51CC2C71B5A CRC64;

Query Match 2.1%; Score 11; DB 10; Length 105;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGKAIDV 442
DB 51 KVKGKAIDV 61

RESULT 4

Q8MFR8 PRELIMINARY; PRT; 588 AA.
AC Q8MFR8; PRELIMINARY; PRT; 588 AA.
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Plastidic ATP/ADP transporter (Fragment).
OS Citrus hybrid cultivar.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=171250;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Weiss D., Goldschmidt E.E.;
RT "Citrus mRNA for plastidic ATP/ADP transporter.";

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY098893; AAM29152.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Chloroplast.
 FT NON_TER
 SQ SEQUENCE 588 AA; 64735 MW; B4B6575BB36337B CRC64;

Query Match 2.1%; Score 11; DB 8; Length 588;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 VKGKAIDVV 442
 |||||
 DB 483 VKGKAIDVV 493

RESULT 5

092HV4 PRELIMINARY; PRT; 511 AA.
 AC 092HV4;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE ADP_ATP carrier protein.
 GN T1C4 OR RC0666.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.",
 RL Science 293:2093-2098(2001).
 DR EMBL; AE008626; AAL03204.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Complete proteome.

SQ SEQUENCE 511 AA; 58036 MW; E194D7E21591BE02 CRC64;

Query Match 1.9%; Score 10; DB 16; Length 511;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 LFDSTKEMAY 424
 |||||
 DB 422 LFDSTKEMAY 431

RESULT 6

08KPN2 PRELIMINARY; PRT; 514 AA.
 AC 08KPN2;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Non-mitochondrial nucleotide transport protein.
 GN NNT.
 OS Holospora obtusa.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Holosporaceae; Holospora.
 OX NCBI_TaxID=49893;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Linka N., Hurka H., Jang F.B., Burger G., Winkler H.H., Stamme C.,
 RA Urbany C., Seil I., Kuech J., Neuhaus E.H.;

RT "Phylogenetic relationships of non-mitochondrial nucleotide transport
 RT proteins in bacteria and eukaryotes."
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY120885; AAM80566.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 514 AA; 57531 MW; B6C3E24B9289E82F CRC64;

Query Match 1.9%; Score 10; DB 2; Length 514;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 TREMAVYPLD 428
 |||||
 DB 423 TREMAVYPLD 432

RESULT 7

09AS91 PRELIMINARY; PRT; 644 AA.
 AC 09AS91;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Putative plastidic ATP/ADP-transporter.
 GN P0707D10.36 OR P0038D11.15.
 OS Oryza sativa (Rice), and
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriophytaceae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530, 39947;
 RN [1]
 RP SEQUENCE FROM N.A.

RC SPECIES=O.sativa; STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0707D10.36";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0038D11.15";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP002910; BAB40979.1; -
 DR EMBL; AP003234; BAB05539.1; -
 DR Gramene; 09AS91; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; T1C; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 SQ SEQUENCE 644 AA; 68960 MW; B3531863E48334NC CRC64;

Query Match 1.9%; Score 10; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 0.82;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 433 VKGKAIDVV 442
 |||||
 DB 526 VKGKAIDVV 535

RESULT 8

P95024 PRELIMINARY; PRT; 463 AA.
 AC P95024;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hypothetical protein RV2529.
 GN RV2529 OR MTCY159.27C.

OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544 (1998).
 DR EMBL; Z83663; CAB06179.1; -
 DR Tuberculex; RV2529; -
 DR InterPro; IPR006166; ERCC4.
 DR Pfam; PF02732; ERCC4; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 463 AA; 50192 MW; 5AC9AAB26394A97F CRC64;
 QY 437 AADVVAAR 445
 DB 218 AADVVAAR 226
 Query Match 1.7%; Score 9; DB 16; Length 463;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
 OSVUG6 PRELIMINARY; PRT; 463 AA.
 ID OSVUG6;
 AC OSVUG6;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein MT2604.
 GN MT2604
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE007096; AAK46914.1; -
 DR TIGR; MT2604; -
 DR InterPro; IPR006166; ERCC4.
 DR Pfam; PF02732; ERCC4; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 463 AA; 50201 MW; 1A7D6A019458E2B7 CRC64;

Query Match 1.7%; Score 9; DB 16; Length 463;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 AADVVAAR 445
 DB 218 AADVVAAR 226

RESULT 10
 ID Q92J16 PRELIMINARY; PRT; 498 AA.
 AC Q92J16;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ADP-ATP carrier protein.
 GN TLCL OR RC0081.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renseto-Audiffren P., Fournier P.-E., Barbe V.,
 Saeon D., Roux V., Cozzari P., Weissenbach J., Claverie J.-M.,
 Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098 (2001).
 DR EMBL; AE008575; AL02619.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Complete proteome.
 SQ SEQUENCE 498 AA; 56661 MW; 97B5EB9C5BCC56B CRC64;
 QY 246 NKNVLTDP 254
 DB 241 NKNVLTDP 249
 Query Match 1.7%; Score 9; DB 16; Length 498;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
 OSVH67 PRELIMINARY; PRT; 671 AA.
 ID OSVH67;
 AC OSVH67;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Plastidic ATP/ADP transporter (Fragment).
 GN ATP1.
 OS Galdieria sulphuraria.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Galdieria.
 OX NCBI_TaxID=130081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stehme C., Neuhäus E.H.;
 RT "Analysis of the plastidic ATP/ADP transporter from the red algae
 R. Galdieria sulphuraria";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ251356; CAC80882.1; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 DR Chloroplast.
 FT NON TER
 SQ SEQUENCE 671 AA; 73829 MW; 2D25A868F2CFFD7 CRC64;

Query Match 1.7%; Score 9; DB 8; Length 671;
 Best Local Similarity 100.0%; Pred. No. 8.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 GTPLMLAV 398

Db 534 GTTPMLAV 542

RESULT 12

Q9YFB6 PRELIMINARY; PRT; 111 AA.
 AC Q9YFB6
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE0325.
 GN APE0325.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankal A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Ogunchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000059; BAA79280.1;
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 111 AA; 10742 MW; CP9180F4DF1A27B8 CRC64;

Query Match 1.6%; Score 8; DB 17; Length 111;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVNGAI 403

Db 3 LAVVNGAI 10

RESULT 13

Q8NFU3 PRELIMINARY; PRT; 115 AA.
 AC Q8NFU3
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE KAT protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wenzel K., Hohe M.R., Bauer D.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF394442; AAM69839.1;
 DR InterPro; IPR001763; Rhodanese-like.
 DR Pfam; PF00581; Rhodanese; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PSS0206; RHODANES; 1.
 SQ SEQUENCE 115 AA; 12530 MW; 01963F5B8186B3A2 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 115;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202

Db 14 SLASGRA 21

RESULT 14

Q9BV22 PRELIMINARY; PRT; 122 AA.
 AC Q9BV22
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC tissue=Lymph;
 RA Strausberg R.,
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC001699; AAH01699.1;
 DR InterPro; IPR001763; Rhodanese-like.
 DR Pfam; PF00581; Rhodanese; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PSS0206; RHODANES; 1.
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 122 AA; 13187 MW; 0C63813B165ADE06 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 122;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202

Db 27 SLASGRA 34

RESULT 15

O35362 PRELIMINARY; PRT; 154 AA.
 AC O35362
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Casr-rs1 (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98008915; PubMed=9344650;
 RA Hinson T.K., Damodaran T.V., Chen J., Zhang X., Qumaiyeh M.B.,
 RA Seldin M.F., Quarles L.D.;
 RT "Identification of putative transmembrane receptor sequences
 homologous to the calcium-sensing G-protein-coupled receptor."
 RL Genomics 45:279-289(1997).
 DR EMBL; AF022250; AAB83974.1;
 DR InterPro; IPR001091; GPCR_MGR.
 DR InterPro; IPR000337; GPCR_MGR.
 DR InterPro; IPR004073; Vmron_receptor2.
 DR Pfam; PF00003; 7tm_3; 1.
 DR PRINTS; PR00248; GPCR_MGR.
 DR PRINTS; PR01535; VOMERONASL2R.
 DR PROSITE; PSS0259; G_PROTEIN_RECP_F3_4; 1.
 DR PROSITE; PSS00093; N4_MTASE; 1.
 FT NON_TER
 SQ SEQUENCE 154 AA; 16875 MW; 325F445080A1CB81 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 154;
 Best Local Similarity 100.0%; Pred. No. 24;


```

RT "DEFOG: A Practical Scheme for Deciphering Families of Genes.";
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF399631; AAK5116.1; -.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KW Receptor.
FT NON_TER 1
FT NON_TER 217
SQ SEQUENCE 217 AA; 24112 MW; D0855DEDE6F09384 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 4; Length 217;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101
Db 209 LFYAVGTP 216

RESULT 20
Q8IU63 PRELIMINARY; PRT; 238 AA.
ID Q8IU63
AC Q8IU63
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Olfactory receptor.
GN 6M1-16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Voiz A., Ehlers A., Younger R., Forbes S., Trowsdale J., Beck S.,
RA Ziegler A.;
RT "Complex transcriptional control of MHC-linked olfactory receptor
RT genes includes long distance and extensive alternative splicing; exon
RT sharing and premature polyadenylation.";
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ459845; CAD31037.1; -.
DR EMBL; AJ459846; CAD31038.1; -.
KW Receptor.
SQ SEQUENCE 238 AA; 26550 MW; B9BB34DBBF6328EF CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 4; Length 238;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 LFYAVGTP 101
Db 196 LFYAVGTP 203

RESULT 21
Q95J13 PRELIMINARY; PRT; 245 AA.
ID Q95J13
AC Q95J13
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Angiotensinogen (Fragment).
GN REN.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OC NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=504, and 505;
RA Sarta Y.;
RT "Comparison of DNA and protein polymorphisms between humans and
RT chimpanzees.";
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
CC -1-SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AB062027; BAB55856.1; -.
DR EMBL; AB062028; BAB55857.1; -.
DR InterPro; IPR000227; Angiotensngn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR PRINTS; PR00654; ANGIOTENNGN.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
FT NON_TER 245
FT NON_TER 245
SQ SEQUENCE 245 AA; 26317 MW; E0092390B9803E0B CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 6; Length 245;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 ADRQATL 133
Db 155 ADRQATL 162

RESULT 22
Q85089 PRELIMINARY; PRT; 247 AA.
ID Q85089
AC Q85089
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Env protein (Fragment).
GN ENV.
OS Primate T-lymphotropic virus 3.
OC Viruses; Retrod viruses; Retroviridae; Deltaretrovirus.
OC NCBI_TaxID=194443;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STLV-PH 969;
RX MEDLINE=94195839; PubMed=7908445;
RA Goubau P., Van Brussel M., Vandamme A., Liu H.F., Desmyter J.;
RT "A primate T-lymphotropic virus, PTLV-L, different from human T-
RT lymphotropic viruses type I and II, in a wild-caught baboon (Papio
RT hamadryas).";
RL Proc. Natl. Acad. Sci. U.S.A. 91:2848-2852(1994).
DR EMBL; Z29673; CA82771.1; -.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 247 AA; 27207 MW; B8EDA1B7B951FE54 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 15; Length 247;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 IAIWLVA 487
Db 73 IAIWLVA 80

RESULT 23
Q9YFU1 PRELIMINARY; PRT; 282 AA.
ID Q9YFU1
AC Q9YFU1
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 282AA long hypothetical bacteriochlorophyll synthase.
GN APE0159.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcates;

```

```

OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=9910339; PubMed=10382966;
RA Kawabatayaai Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kusuda N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RT DNA Res. 6:83-101(1999).
DR EMBL; AP000058; BAA79070.1; -.
DR InterPro; IPR000812; TFIIB_euk.
DR InterPro; IPR000537; UblA.
DR Pfam; PF01040; UblA; 1.
DR PROSITE; PS00782; TFIIB; 1.
KW Complete proteome.
SQ SEQUENCE 282 AA; 29648 MW; D39C1EAD0B9E600 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 17; Length 282;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLTLTLV 294
DB 234 ILLTLTLV 241

RESULT 24
Q8K125 PRELIMINARY; PRT; 294 AA.
AC Q8K125;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034887; AAH34887.1; -.
DR InterPro; IPR003599; IG_1-like.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG_LIKE; 2.
KW Hypothetical protein.
SQ SEQUENCE 294 AA; 32368 MW; 6DBBE836659735F2 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 11; Length 294;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGGLTMA 240
DB 280 VSGGLTMA 287

RESULT 25
Q8VF92 PRELIMINARY; PRT; 306 AA.
ID Q8VF92

```

```

AC Q8VF92;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Olfactory receptor MOR256-29.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang X., Firestein S.J.;
RT "The olfactory receptor gene superfamily of the mouse.";
RL Nat. Neurosci. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Adams M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY073642; AAL61305.1; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECCEP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECCEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 306 AA; 34054 MW; 361F56047B31E9CA CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 11; Length 306;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

RESULT 26
Q9WV14 PRELIMINARY; PRT; 310 AA.
AC Q9WV14;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 573K1.2 (Mm17M1-3 (Novel 7 transmembrane receptor (Rhodopsin family)
DE (Olfactory receptor like) protein)).
DE (Olfactory receptor like) protein)).
GN GABBR1 OR 573K1.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44994.1; -.
DR MGD; MGI:1860139; Gabbr1.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECCEP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECCEP_F1_2; 1.
KW Receptor; Transmembrane.
SQ SEQUENCE 310 AA; 34540 MW; 8C069BF46E35DEAD CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 11; Length 310;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

```

RESULT 27

08VF82 PRELIMINARY; PRT; 310 AA.
AC 08VF82;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Olfactory receptor MOR256-21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang X., Firestein S.J.;
RT "The olfactory receptor gene superfamily of the mouse."
RL Nat. Neurosci. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Adams M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY073621; AAL61284.1; -.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM Receptor.
SQ SEQUENCE 310 AA; 34641 MW; 0C8572833AD7A66 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 310;
Best Local Similarity 100.0%; Pred. No. 43;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||
Db 274 LFYAVGTP 281

RESULT 28

09WV13 PRELIMINARY; PRT; 312 AA.
ID 09WV13;
AC 09WV13;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.4 (Mm17M1-1 (Novel 7 transmembrane receptor (Rhodopsin family)
(Olfactory receptor like) protein)).
GN GABBR1 OR 573K1.4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44996.1; -.
DR MGD; MGI:1860139; Gabbr1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34988 MW; 630498DA34FCD1D0 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||

Db 274 LFYAVGTP 281

RESULT 29

09WV11 PRELIMINARY; PRT; 312 AA.
ID 09WV11;
AC 09WV11;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.8 (Mm17M1-2 (Novel 7 transmembrane receptor (rhodopsin family)
(Olfactory receptor like) protein)).
GN GABBR1 OR 573K1.8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB45000.1; -.
DR MGD; MGI:1860139; Gabbr1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34786 MW; DAA7507C77110AD8 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||
Db 274 LFYAVGTP 281

RESULT 30

09R022 PRELIMINARY; PRT; 312 AA.
ID 09R022;
AC 09R022;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 573K1.3 (Mm17M1-4 (Novel 7 transmembrane receptor (rhodopsin family)
(Olfactory receptor like) protein)).
GN GABBR1 OR 573K1.3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Younger R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078630; CAB44995.1; -.
DR MGD; MGI:1860139; Gabbr1.
DR InterPro; IPR000276; GPCR_Rhodpn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECPE_FL_2; 1.
KM Receptor; Transmembrane.
SQ SEQUENCE 312 AA; 34786 MW; DAA7507C77110AD8 CRC64;

Query Match 1.6%; Score 8; DB 11; Length 312;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
|||||

```

DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
DR RECEPTOR; Transmembrane.
SQ SEQUENCE 312 AA; 3499 MW; 83374B1A4C6F912D CRC64;

Query March
Best Local Similarity 1.6%; Score 8; DB 11; Length 312;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFVAVGTP 101
Db 274 LFVAVGTP 281

RESULT 31
Q8F697 PRELIMINARY; PRT; 340 AA.
AC Q8F697;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN IAI412.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
CX NCBI_TaxID=173;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar Iai;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011321; AAA48611.1.-.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 340 AA; 38370 MW; 1A0520FA1B5DC0F CRC64;

Query March
Best Local Similarity 1.6%; Score 8; DB 16; Length 340;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 LGLVAIL 144
Db 318 LGLVAIL 325

RESULT 32
Q9P042 PRELIMINARY; PRT; 342 AA.
AC Q9P042;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE HSC108.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
MEDLINE=20499367; PubMed=11042152;
RA Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Fu G.,
RA Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560 (2000).
DR EMBL; AF161458; AAF29073.1.-.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR001972; Stomatin.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.

```

```

DR SMART; SMO0244; PHB; 1.
SQ SEQUENCE 342 AA; 37145 MW; 7958C0B3BDE53E5 CRC64;

Query March
Best Local Similarity 1.6%; Score 8; DB 4; Length 342;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202
Db 4 SLASGRA 11

RESULT 33
Q9UZ1 PRELIMINARY; PRT; 356 AA.
AC Q9UZ1;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Membrane associated protein SLP-2 (Stomatin-like protein 2) (Stomatin-
DE like 2) (Hypothetical protein FLJ14499).
GN HUSLP2 OR SLP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart muscle;
RA Wang Y., Morrow J.S.;
RT "Identification and Characterization of Human SLP-2, a Novel Homologue
RT of Stomatin (band 7.2b) Present in Erythrocytes and Other Tissues.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Owczarek C.M., Treutlein H.R., Portbury K.J., Gulluyan L.M., Kola I.,
RA Hertzog P.J.;
RT "A novel member of the stomatin/EPB72/MCC-2 family, stomatin-like
RT protein 2 (SLP2), is ubiquitously expressed and localizes to HSA
RL chromosome 9p13.1.";
RN (3)
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Minomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, Lung, and Skin;
RA Straubeberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190167; AAF09142.1.-.
DR EMBL; AF282596; AAF91466.1.-.
DR EMBL; BC003025; AAH03025.1.-.
DR EMBL; BC002442; AAH02442.1.-.
DR EMBL; AK027405; BAB5091.1.-.
DR EMBL; BC014990; AAH14990.1.-.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR001972; Stomatin.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SMO0244; PHB; 1.
KW Hypothetical protein.
SQ SEQUENCE 356 AA; 38534 MW; 672331B57C82654E CRC64;

Query March
Best Local Similarity 1.6%; Score 8; DB 4; Length 356;

```


Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202

Db 17 SLASGRA 24

RESULT 34

Q96FY2 PRELIMINARY; PRT; 356 AA.
AC Q96FY2; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Stomatol-like 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010152; AAH10152.1; -
DR InterPro; IPR001107; Band_7;
DR InterPro; IPR001972; Stomatol.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
SQ SEQUENCE 356 AA; 38518 MW; 6FC0BF899E1919A7 CRC64;

Query Match 1.6%; Score 8; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202

Db 17 SLASGRA 24

RESULT 35

060376 PRELIMINARY; PRT; 357 AA.
AC 060376; 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast or foreskin.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stiliwgen S.,
RA Phan H., Velasco N., Garmes J., Dangnan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Altix C., Andreise T.,
RA Frankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Olsen A.O.,
RA Montgomey M., Ow D., Nolan M., Triong S., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a human P1 clone containing the XRC9 DNA repair
gene";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004472; AAC07983.1; -
DR Genew; HGNC14559; STOML2.
DR InterPro; IPR001107; Band_7;
DR InterPro; IPR001972; Stomatol.
DR Pfam; PF01145; Band_7; 1.

DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
SQ SEQUENCE 357 AA; 38749 MW; 512632B83028135A CRC64;

Query Match 1.6%; Score 8; DB 4; Length 357;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 SLASGRA 202

Db 4 SLASGRA 11

RESULT 36

Q9SIR6 PRELIMINARY; PRT; 360 AA.
AC Q9SIR6; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE At2g25270 protein.
GN AT2G25270.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Puji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Beell C.R., Kechum K.A., Lee J.-J., Rouning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gall J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007070; AAD23664.1; -
SQ SEQUENCE 360 AA; 39992 MW; EAPD5D2A9E57870F CRC64;

Query Match 1.6%; Score 8; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 VMLSLMF 169

Db 325 VMLSLMF 332

RESULT 37

08GLH2 PRELIMINARY; PRT; 400 AA.
AC 08GLH2; 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Acetate kinase (EC 2.7.2.1).
GN ACK.
OS Photobacterium temperata.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_Taxid=171441;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=K122;
RA Joyce S.A., Clarke D.J.;
RT "A hexa homolog from Phototrichobdus regulates pathogenicity, symbiosis
RT and phenotypic variation.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY137386; AAN08359.1; -.
KW Kinase, Transferrase.
SQ SEQUENCE 400 AA; 43388 MW; AD4D6555618F1DF CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 400;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AARFGKSG 450
Db 363 AARFGKSG 370

RESULT 38
Q8YV45 PRELIMINARY; PRT; 480 AA.
AC Q8YV45;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Hypothetical protein A111007.
GN A111007.
OS Arabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Arabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003584; BAB72964.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 480 AA; 54939 MW; 22908DF99BBC417A CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 16; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 288 LLTLTLVI 295
Db 429 LLTLTLVI 436

RESULT 39
Q9GLN8 PRELIMINARY; PRT; 485 AA.
AC Q9GLN8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Angiotensinogen.
GN AGT.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20469400; PubMed=11013071;
RA Dufour C., Casane D., Denton D., Wickings J., Corvol P.,
RA Jeunemaitre X.;
RT "Human-Chimpanzee DNA sequence variation in the four major genes of

```

```

RT the renin angiotensin system.";
RL Genomics 69:14-26(2000).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF193461; AAG30306.1; -.
DR EMBL; AF193458; AAG30306.1; JOINED.
DR EMBL; AF193459; AAG30306.1; JOINED.
DR EMBL; AF193460; AAG30306.1; JOINED.
DR InterPro; IPR000227; Angiotensngn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR PRINTS; PR00654; ANGIOTENNGN.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
SQ SEQUENCE 485 AA; 53110 MW; C14C67B49A33F05F CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 6; Length 485;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 ADRLOAIL 133
Db 155 ADRLOAIL 162

RESULT 40
Q9GLP7 PRELIMINARY; PRT; 485 AA.
AC Q9GLP7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Angiotensinogen.
GN AGT.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX Shattuck-Eidens D., McGrath M., Stone S.;
RT "Germline mutations in the angiotensinogen gene cause predisposition
RT to type 1 diabetes mellitus.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF188487; AAG29056.1; -.
DR InterPro; IPR000227; Angiotensngn.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR PRINTS; PR00654; ANGIOTENNGN.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
SQ SEQUENCE 485 AA; 53140 MW; 49EFB54FAF31F8ADC CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 6; Length 485;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 ADRLOAIL 133
Db 155 ADRLOAIL 162

Search completed: November 25, 2003, 10:07:00
Job time : 70 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:05:06 ; Search time 23 Seconds
(without alignments)
947.395 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515

Sequence: 1 MKTEKPKPKLRSLFMPH.....AQSALKEQVAVQSDAPASS 515

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	313	60.8	524	US-09-198-452A-369	Sequence 369, App
2	8	1.6	111	US-09-328-352-5992	Sequence 5992, Ap
3	8	1.6	137	US-08-974-546-3	Sequence 3, Appl
4	8	1.6	172	US-09-252-991A-23710	Sequence 23710, A
5	8	1.6	293	US-09-509-347-7	Patent No. 5189147
6	8	1.6	303	US-09-509-347-7	Sequence 7, Appl
7	8	1.6	485	US-09-384-212-2	Sequence 2, Appl
8	8	1.6	1013	US-09-415-522-8	Sequence 8, Appl
9	8	1.6	1964	US-09-467-997-1	Sequence 1, Appl
10	8	1.4	16	US-07-894-063A-1	Sequence 1, Appl
11	7	1.4	16	US-07-894-063A-2	Sequence 2, Appl
12	7	1.4	77	US-09-393-634-11	Sequence 11, Appl
13	7	1.4	110	US-09-732-210-1020	Sequence 1020, Ap
14	7	1.4	117	US-08-444-818-44	Sequence 44, Appl
15	7	1.4	132	US-08-444-818-52	Sequence 52, Appl
16	7	1.4	159	US-09-252-991A-29736	Sequence 29736, A
17	7	1.4	193	US-08-308-821A-6	Sequence 6, Appl
18	7	1.4	193	US-08-311-627-6	Sequence 6, Appl
19	7	1.4	193	US-08-765-381-15	Sequence 15, Appl
20	7	1.4	197	US-08-804-372A-32	Sequence 32, Appl
21	7	1.4	211	PCT-US94-04174-18	Sequence 18, Appl
22	7	1.4	216	US-09-252-991A-21062	Sequence 21062, A
23	7	1.4	218	US-09-328-352-4385	Sequence 4385, Ap
24	7	1.4	225	US-09-107-532A-7302	Sequence 7302, Ap
25	7	1.4	253	US-09-333-599-2	Sequence 2, Appl
26	7	1.4	253	US-09-499-781-2	Sequence 2, Appl
27	7	1.4	259	US-09-252-991A-19115	Sequence 19115, A

28	7	1.4	279	US-09-328-352-7144	Sequence 7144, Ap
29	7	1.4	293	US-09-414-276-3	Sequence 3, Appl
30	7	1.4	317	US-09-134-001C-4537	Sequence 4537, Ap
31	7	1.4	335	US-09-414-276-6	Sequence 6, Appl
32	7	1.4	357	US-09-107-532A-5132	Sequence 5132, Ap
33	7	1.4	362	US-09-252-991A-18068	Sequence 18068, A
34	7	1.4	375	US-09-622-439-2	Sequence 2, Appl
35	7	1.4	414	US-09-107-532A-6371	Sequence 6371, Ap
36	7	1.4	423	US-09-328-352-6097	Sequence 6097, Ap
37	7	1.4	425	US-09-328-352-7964	Sequence 7964, Ap
38	7	1.4	452	US-09-252-991A-20578	Sequence 20578, A
39	7	1.4	504	US-07-853-985A-6	Sequence 6, Appl
40	7	1.4	504	US-07-681-703B-6	Sequence 6, Appl
41	7	1.4	504	US-08-184-236-6	Sequence 6, Appl
42	7	1.4	504	US-08-407-410B-6	Sequence 6, Appl
43	7	1.4	504	US-08-485-500-6	Sequence 6, Appl
44	7	1.4	504	PCT-US91-02370-6	Sequence 6, Appl
45	7	1.4	504	PCT-US94-04174-6	Sequence 6, Appl
46	7	1.4	532	US-09-252-991A-27288	Sequence 27288, A
47	7	1.4	541	US-08-867-611-8	Sequence 8, Appl
48	7	1.4	541	PCT-US92-06965A-13	Sequence 13, Appl
49	7	1.4	574	US-08-309-512-9	Sequence 9, Appl
50	7	1.4	574	PCT-US92-08766A-9	Sequence 9, Appl
51	7	1.4	579	US-09-173-151A-2	Sequence 2, Appl
52	7	1.4	590	US-09-208-140-4	Sequence 4, Appl
53	7	1.4	591	US-09-208-140-2	Sequence 2, Appl
54	7	1.4	593	US-08-722-806A-2	Sequence 2, Appl
55	7	1.4	593	US-09-337-028-2	Sequence 2, Appl
56	7	1.4	593	US-09-597-877-2	Sequence 2, Appl
57	7	1.4	604	US-09-345-473B-17	Sequence 17, Appl
58	7	1.4	686	US-09-614-259-33	Sequence 33, Appl
59	7	1.4	746	US-09-252-991A-18587	Sequence 18587, A
60	7	1.4	785	US-09-252-991A-32952	Sequence 32952, A
61	7	1.4	798	US-08-867-611-36	Sequence 36, Appl
62	7	1.4	919	US-09-437-054A-17	Sequence 17, Appl
63	7	1.4	1011	US-08-850-328-5	Sequence 5, Appl
64	7	1.4	1786	US-08-444-818-54	Sequence 54, Appl
65	7	1.4	2261	US-08-444-818-65	Sequence 65, Appl
66	7	1.4	2436	US-08-444-818-75	Sequence 75, Appl
67	7	1.4	2772	US-08-444-818-89	Sequence 89, Appl
68	7	1.4	2955	US-08-442-805A-3	Sequence 3, Appl
69	7	1.4	2955	US-08-442-805A-3	Sequence 3, Appl
70	7	1.4	2955	US-08-443-900A-3	Sequence 3, Appl
71	7	1.4	2955	US-08-444-818-124	Sequence 124, App
72	7	1.4	2955	US-08-249-843-3	Sequence 3, Appl
73	7	1.4	2995	US-08-444-818-138	Sequence 138, App
74	7	1.4	3011	US-08-188-281B-1	Sequence 1, Appl
75	7	1.4	3011	US-08-453-552-1	Sequence 1, Appl
76	7	1.4	3011	US-08-440-103-36	Sequence 36, Appl
77	7	1.4	3011	US-08-440-103-36	Sequence 36, Appl
78	7	1.4	3011	US-08-440-542-36	Sequence 36, Appl
79	7	1.4	3011	US-07-910-760-10	Sequence 10, Appl
80	7	1.4	3011	US-08-440-519-10	Sequence 10, Appl
81	7	1.4	3011	US-08-231-368-36	Sequence 36, Appl
82	7	1.4	3011	US-08-440-210-36	Sequence 36, Appl
83	7	1.4	3011	US-08-710-637-1	Sequence 1, Appl
84	7	1.4	3011	US-08-813-678A-6	Sequence 6, Appl
85	7	1.4	3011	US-08-811-566-20	Sequence 20, Appl
86	7	1.4	3011	US-08-444-818-177	Sequence 177, App
87	7	1.4	3011	US-09-014-416-1	Sequence 1, Appl
88	7	1.4	3011	US-09-014-416-5	Sequence 5, Appl
89	7	1.4	3011	US-08-529-169A-6	Sequence 6, Appl
90	7	1.4	3011	US-09-388-874-2	Sequence 2, Appl
91	7	1.4	3011	US-09-046-604-36	Sequence 36, Appl
92	7	1.4	3011	US-08-440-549-10	Sequence 10, Appl
93	7	1.4	3011	US-08-850-328-1	Sequence 1, Appl
94	7	1.4	3011	US-09-034-756-20	Sequence 20, Appl
95	7	1.4	3011	US-09-483-799-6	Sequence 6, Appl
96	7	1.4	3011	US-09-916-359-2	Sequence 2, Appl
97	7	1.4	3011	PCT-US91-02225-10	Sequence 10, Appl
98	7	1.4	3011	PCT-US93-00907-1	Sequence 1, Appl
99	7	1.4	3011	PCT-US94-07280-1	Sequence 1, Appl
100	7	1.4	3012	PCT-US95-01087-1	Sequence 1, Appl
				US-08-811-566-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-198-452A-369
Sequence 369, Application US/09198452A
Patent No. 6559394
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 369
LENGTH: 524
TYPE: PR
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-369

Query Match 60.8%; Score 313; DB 4; Length 524;
Best Local Similarity 99.6%; Pred. No. 3,982,399;
Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKEEKPFGKLSFLPHTHEKLVLMFMFCTTNTVLRDTKDTLIVGAPSSGA 60
DB 10 MTKEEKPFGKLSFLPHTHEKLVLMFMFCTTNTVLRDTKDTLIVGAPSSGA 69
QY 61 EALPFIKFWLVVPCALIFMLIVAKLSNLSKQALFYVGPPELIFPLPVTIYPLADVL 120
DB 70 EALPFIKFWLVVPCALIFMLIVAKLSNLSKQALFYVGPPELIFPLPVTIYPLADVL 129
QY 121 HPTFEADRLQALIPPGILGLVAIIRNNTFAFYVLAELMGSVMSLMFNGFANEITKHE 180
DB 130 HPTFEADRLQALIPPGILGLVAIIRNNTFAFYVLAELMGSVMSLMFNGFANEITKHE 189
QY 181 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTVSGVLVMA 240
DB 190 AKRFYALFGIGANISLASGRAIYVASKLRASVSEGVDPWGISIRLIMAMTVSGVLVMA 249
QY 241 SYMMINNVLTDRFYVPEEMQKKGAKPRMNKDSFVLDRSPYLLLTLLVIAVIGIC 300
DB 250 SYMMINNVLTDRFYVPEEMQKKGAKPRMNKDSFVLDRSPYLLLTLLVIAVIGIC 309
QY 301 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFGVNVIRKFGMLTGALV 360
DB 310 INLEIETWKSQKLQYNNNDYSEFMGNFSFWTGVSVLIMLFGVNVIRKFGMLTGALV 369
QY 361 TPVWVLLTGIIVPALVIFRNQASGLVAMFGTTPMLAVVVGALQNLISKSTKVALFDSTK 420
DB 370 TPVWVLLTGIIVPALVIFRNQASGLVAMFGTTPMLAVVVGALQNLISKSTKVALFDSTK 429
QY 421 EMAYIPIDQOKYKGAIVVVARPKSGALIQOGLVYCGSIGMTPEYLVILFFI 480
DB 430 EMAYIPIDQOKYKGAIVVVARPKSGALIQOGLVYCGSIGMTPEYLVILFFI 489
QY 481 AIMVLSATKLNKLFALQSAKCEQVADSDAPASS 515
DB 490 AIMVLSATKLNKLFALQSAKCEQVADSDAPASS 524

RESULT 2

US-09-328-352-5992
Sequence 5992, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5992
LENGTH: 111
TYPE: PR
ORGANISM: Acinetobacter baumannii
US-09-328-352-5992

Query Match 1.6%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 5,5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 IVSGVLVM 239
DB 98 IVSGVLVM 105

RESULT 3

US-08-974-546-3
Sequence 3, Application US/08974546
Patent No. 5945287
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,546
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0428
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT18
CLONE: 3172266
US-08-974-546-3

Query Match 1.6%; Score 8; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 6,8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLLASGRA 202
DB 36 SLLASGRA 43

```
RESULT 4
US-09-252-991A-23710
; Sequence 23710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23710
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23710

Query Match
1.6%; Score 8; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 LPPGLGL 140
DB 142 LPPGLGL 149

RESULT 5
5189147-3
; Patent No. 5189147
; APPLICANT: SAITO, HARUO; KRANZ, DAVID M.; EISEN, HERMAN N.;
; TONEGAWA, SUSUMU
; TITLE OF INVENTION: METEODIMERIC T LYMPHOCYTE RECEPTOR
; ANTIBODY
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/271,216
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 666,988
; FILING DATE: 31-OCT-1984
; APPLICATION NUMBER: 620,122
; FILING DATE: 13-JUN-1984
; SEQ ID NO:3
; LENGTH: 293
5189147-3

Query Match
1.6%; Score 8; DB 6; Length 293;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
DB 279 VSGLVMA 286

RESULT 6
US-09-509-347-7
; Sequence 7, Application US/09509347
; Patent No. 6511830
; GENERAL INFORMATION:
; APPLICANT: KYOMA HAKKO KYOGYO CO., LTD.
; TITLE OF INVENTION: A KILLER T CELL RECEPTOR RECOGNIZING HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS
; FILE REFERENCE: PH-585PCT
; CURRENT APPLICATION NUMBER: US/09/509,347
; CURRENT FILING DATE: 2000-03-27
```

```
; PRIOR APPLICATION NUMBER: JP97/262536
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-509-347-7

Query Match
1.6%; Score 8; DB 4; Length 303;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVMA 240
DB 289 VSGLVMA 296

RESULT 7
US-09-384-212-2
; Sequence 2, Application US/09384212
; Patent No. 6177252
; GENERAL INFORMATION:
; APPLICANT: Lalouel, Jean-Marc
; APPLICANT: Rohrwasser, Andreas
; TITLE OF INVENTION: Method to Determine Predisposition to Hypertension
; FILE REFERENCE: 2323-142
; CURRENT APPLICATION NUMBER: US/09/384,212
; CURRENT FILING DATE: 1999-08-27
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-384-212-2

Query Match
1.6%; Score 8; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ADRLOAIL 133
DB 155 ADRLOAIL 162

RESULT 8
US-09-415-522-8
; Sequence 8, Application US/09415522A
; Patent No. 6291660
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas
; APPLICANT: Wendland, Juergen
; APPLICANT: Philippsen, Peter
; TITLE OF INVENTION: No. 6291660el Fungal Genes Required For No. 6291660mal Growth And
; FILE REFERENCE: CGC2046
; CURRENT APPLICATION NUMBER: US/09/415,522A
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-415-522-8

Query Match
1.6%; Score 8; DB 3; Length 1013;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASGR 201
```

Db 79 ISLASGR 86

RESULT 9

US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitejowski, Jan
; APPLICANT: Kitejowski, Hendrik
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PT-US
; CURRENT APPLICATION NUMBER: US/09/467,997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1964
; TYPE: PRT
; ORGANISM: mouse
US-09-467-997-1

Query Match 1.6%; Score 8; DB 4; Length 1964;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 GTTPLMLA 397
Db 1696 GTTPLMLA 1703

RESULT 10

US-07-894-063A-1
; Sequence 1, Application US/07894063A
; Patent No. 5980899
; GENERAL INFORMATION:
; APPLICANT: BERZOFISKY, Jay A.
; APPLICANT: SHIRAI, Mutsunori
; APPLICANT: AKATSUKA, Toshitaka
; APPLICANT: FEINSTONE, Stephen M.
; TITLE OF INVENTION: PEPTIDE FOR STIMULATION OF CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTES SPECIFIC FOR HEPATITIS C VIRUS IN A MAMMAL
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,063A
; FILING DATE: 19920610
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/162/NHND
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

US-07-894-063A-1

Query Match 1.4%; Score 7; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
Db 6 TGAIVTP 12

RESULT 11

US-07-894-063A-2
; Sequence 2, Application US/07894063A
; Patent No. 5980899
; GENERAL INFORMATION:
; APPLICANT: BERZOFISKY, Jay A.
; APPLICANT: SHIRAI, Mutsunori
; APPLICANT: AKATSUKA, Toshitaka
; APPLICANT: FEINSTONE, Stephen M.
; TITLE OF INVENTION: PEPTIDE FOR STIMULATION OF CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTES SPECIFIC FOR HEPATITIS C VIRUS IN A MAMMAL
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,063A
; FILING DATE: 19920610
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/162/NHND
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

US-07-894-063A-2

Query Match 1.4%; Score 7; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
Db 6 TGAIVTP 12

RESULT 12

US-09-393-634-11
; Sequence 11, Application US/09393634
; Patent No. 6558910
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark

APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. 6558910el Family of Taste Receptors
FILE REFERENCE: 02307E-098000S
CURRENT APPLICATION NUMBER: US/09/393,634
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 77
TYPE: PRT
ORGANISM: Mus sp.
FEATURE:
OTHER INFORMATION: mouse GR01
NAME/KEY: MOD_RES
LOCATION: (1)-(77)
OTHER INFORMATION: Xaa = any amino acid
US-09-393-634-11

Query Match 1.4%; Score 7; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 289 LITLTVI 295
DB 37 LITLTVI 43

RESULT 13
US-09-732-210-1020
Sequence 1020, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunkers, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Mitland, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 1020
LENGTH: 110
TYPE: PRT
ORGANISM: Melhanococcus jannaschii
US-09-732-210-1020

Query Match 1.4%; Score 7; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 82 YAKLSNI 88
DB 58 YAKLSNI 64

RESULT 14
US-08-444-818-44
Sequence 44, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-44

Query Match 1.4%; Score 7; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 356 TGAIVTP 362
DB 80 TGAIVTP 86

RESULT 15
US-08-444-818-52
Sequence 52, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 359-3876
TELEFAX: (508) 359-3885
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-52

Query Match 1.4%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TGAIVTP 362
Db 57 TGAIVTP 63

RESULT 16
US-09-252-991A-29736
Sequence 29736, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29736
LENGTH: 159
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29736

Query Match 1.4%; Score 7; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 PGLIGLV 141
Db 110 PGLIGLV 116

RESULT 17
US-08-308-821A-6
Sequence 6, Application US/08308821A
Patent No. 5885567
GENERAL INFORMATION:
APPLICANT: Sekellick, Margaret J.
APPLICANT: Marcus, Philip I.
APPLICANT: Ferrandino, Anthony F.
TITLE OF INVENTION: AVIAN INTERFERON GENES, NOVEL
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,821A
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,402
FILING DATE: 28-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/139,909
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: UCT93-04A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-308-821A-6

Query Match 1.4%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILTLTL 293
Db 15 ILTLTL 21

RESULT 18
US-08-831-627-6
Sequence 6, Application US/08831627
Patent No. 6020465
GENERAL INFORMATION:
APPLICANT: Sekellick, Margaret J.
APPLICANT: Marcus, Philip I.
APPLICANT: Ferrandino, Anthony F.
TITLE OF INVENTION: CHICKEN INTERFERON GENE AND NOVEL
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,627
FILING DATE: 09-APR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,402
FILING DATE: 28-APR-1994
APPLICATION NUMBER: US 08/139,909
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: UCT93-04A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-831-627-6

Query Match 1.4%; Score 7; DB 3; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 19
US-08-765-381-15
Sequence 15; Application US/08765381
Patent No. 6083724
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
TITLE OF INVENTION: sequences encoding same
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully Scott Murphy and Presser
STREET: 400 Garden City Plaza
CITY: Garden City, New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,381
FILING DATE: 19-DEC-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PNI542/95
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00114
FILING DATE: 05-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESSER, LEOPOLD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-516-742-4343
TELEFAX: 1-516-742-4366
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: avian
IMMEDIATE SOURCE:
CLONE: ChIFN-alpha
US-08-765-381-15

Query Match 1.4%; Score 7; DB 3; Length 193;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 20
US-08-804-372A-32
Sequence 32; Application US/08804372A
Patent No. 6183753
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Wild, Martha A.
APPLICANT: Winslow, Barbara J.
TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,372A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 2552/39115E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 197 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-804-372A-32

Query Match 1.4%; Score 7; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ILLLTLL 293
DB 15 ILLLTLL 21

RESULT 21
PCT-US94-04174-18
Sequence 18; Application PC/TUS9404174
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
TITLE OF INVENTION: HCV Immunodiagnostic Antigens and Antibodies
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter J. Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04174
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-324-0880
TELEFAX: 415-324-0960
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 211 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Clone 36-2 protein sequence
PCT-US94-04174-18

Query Match 1.4%; Score 7; DB 5; Length 211;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 TGAIVTP 362
Db 142 TGAIVTP 148
RESULT 22
US-09-252-991A-21062
Sequence 21062, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/252,991A
FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21062
LENGTH: 216
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21062

Query Match 1.4%; Score 7; DB 4; Length 216;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 QASGLVA 387
Db 113 QASGLVA 119

RESULT 23
US-09-328-352-4385

Sequence 4385, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328,352
FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4385
LENGTH: 218
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4385

Query Match 1.4%; Score 7; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 VILLFTI 480
Db 50 VILLFTI 56

RESULT 24
US-09-107-532A-7302
Sequence 7302, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-8277
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 7302:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...225

SEQUENCE DESCRIPTION: SEQ ID NO: 7302:
US-09-107-532A-7302

Query Match 1.4%; Score 7; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 RDTKDTL 51
Db 76 RDTKDTL 82

RESULT 25
US-09-333-599-2
; Sequence 2, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-333-599-2

Query Match 1.4%; Score 7; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
Db 49 ISLLASG 55

RESULT 26
US-09-499-781-2
; Sequence 2, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/499,781
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/333,599
; PRIOR FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-499-781-2

Query Match 1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
Db 49 ISLLASG 55

RESULT 27

US-09-252-991A-19115
; Sequence 19115, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19115
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19115

Query Match 1.4%; Score 7; DB 4; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 IVPFALV 376
Db 55 IVPFALV 61

RESULT 28
US-09-328-352-7144
; Sequence 7144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7144
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7144

Query Match 1.4%; Score 7; DB 4; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 NISLAS 199
Db 239 NISLAS 245

RESULT 29
US-09-414-276-3
; Sequence 3, Application US/09414276
; Patent No. 6392121
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Palmer, Kenneth
; APPLICANT: Hefferon, Kathleen
; APPLICANT: Mor, Tsafir
; APPLICANT: Antzen, Charles
; TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
; FILE REFERENCE: 4868/84453
; CURRENT APPLICATION NUMBER: US/09/414,276
; CURRENT FILING DATE: 1999-10-07

NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 293
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-09-414-276-3

Query Match 1.4%; Score 7; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPFLIF 106
Db 39 TPFLIF 45

RESULT 30
US-09-134-001C-4537
Sequence 4537, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4537
LENGTH: 317
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4537

Query Match 1.4%; Score 7; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ELKKVLP 29
Db 46 ELKKVLP 52

RESULT 31
US-09-414-276-6
Sequence 6, Application US/09414276
Patent No. 6392121
GENERAL INFORMATION:
APPLICANT: Mason, Hugh
APPLICANT: Palmer, Kenneth
APPLICANT: Hefferon, Kathleen
APPLICANT: Mor, Tsafir
APPLICANT: Arntzen, Charles
TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
FILE REFERENCE: 4868/84453
CURRENT APPLICATION NUMBER: US/09/414,276
CURRENT FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 335
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-09-414-276-6

Query Match 1.4%; Score 7; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPFLIF 106
Db 39 TPFLIF 45

RESULT 32
US-09-107-532A-5132
Sequence 5132, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5132:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...357
SEQUENCE DESCRIPTION: SEQ ID NO: 5132:
US-09-107-532A-5132

Query Match 1.4%; Score 7; DB 4; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 IVEFALV 376
Db 21 IVEFALV 27

RESULT 33
US-09-252-991A-18068
Sequence 18068, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074.788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094.190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18068
LENGTH: 362
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18068

Query Match 1.4%; Score 7; DB 4; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 PFLIFPA 107
DB 177 PFLIFPA 183

RESULT 34
US-09-622-439-2
Sequence 2; Application US/09622439
Patent No. 655344
GENERAL INFORMATION:
APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
TITLE OF INVENTION: A novel G protein coupled receptor protein
FILE REFERENCE: Y9905
CURRENT APPLICATION NUMBER: US/09/622.439
CURRENT FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: JP P1998-060245
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: JP P1999-026774
PRIOR FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 375
TYPE: PRT
ORGANISM: Homo sapiens
US-09-622-439-2

Query Match 1.4%; Score 7; DB 4; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 LAVVGA 402
DB 190 LAVVGA 196

RESULT 35
US-09-107-532A-6371
Sequence 6371, Application US/09107532A
Patent No. 658325
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Walham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107.532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085.598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6371:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...414
SEQUENCE DESCRIPTION: SEQ ID NO: 6371:
US-09-107-532A-6371

Query Match 1.4%; Score 7; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GLVIMAS 241
DB 4 GLVIMAS 10

RESULT 36
US-09-328-352-6097
Sequence 6097, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FIE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6097
LENGTH: 423
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6097

Query Match 1.4%; Score 7; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 ANISLA 198
DB 159 ANISLA 165

RESULT 37
US-09-328-352-7964
Sequence 7964, Application US/09328352
Patent No. 6562958

```

; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7964
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7964

Query Match      1.4%; Score 7; DB 4; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      60 AEAIPFI 66
        |||||
Db      65 AEAIPFI 71

RESULT 38
US-09-252-991A-20578
; Sequence 20578, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20578
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20578

Query Match      1.4%; Score 7; DB 4; Length 452;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      58 SGAEAP 64
        |||||
Db      436 SGAEAP 442

RESULT 39
US-07-853-985A-6
; Sequence 6, Application US/07853985A
; Patent No. 5436318
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungshuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: P.O. BOX 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,985A
; FILING DATE: 19920320
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,611
; FILING DATE: 06-APR-1990
; APPLICATION NUMBER: US 07/594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Dehlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: 4600-0076.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-324-0880
; TELEFAX: 415-324-0960
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 504 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-853-985A-6

Query Match      1.4%; Score 7; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      356 TGAIVTP 362
        |||||
Db      308 TGAIVTP 314

RESULT 40
US-07-681-703B-6
; Sequence 6, Application US/07681703B
; Patent No. 543965
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungshuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/681,703B
; FILING DATE: 05-APR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (415) 324-0880
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-681-703B-6

Query Match 1.4%; Score 7; DB 1; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 TGAIVTP 362 |||||
DB 308 TGAIVTP 314

Search completed: November 25, 2003, 10:08:02
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:07:06 / Search time 55 Seconds
(without alignments)
1727.060 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515
Sequence: 1 MKRTEKPKRGKLRSLFMPH.....AQSALKEQVADSDAPASS 515

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 673684 seqs, 18443283 residues

Word size : 0

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515	100.0	515	9 US-09-892-851-2	Sequence 2, Appl1
2	8	1.6	73	12 US-10-029-386-28700	Sequence 28700, A
3	8	1.6	302	9 US-09-789-697A-21	Sequence 21, Appl1
4	8	1.6	303	12 US-09-795-271-57	Sequence 57, Appl1
5	8	1.6	310	12 US-09-965-422-47	Sequence 47, Appl1
6	8	1.6	312	12 US-10-017-161-508	Sequence 508, App
7	8	1.6	312	12 US-09-965-422-46	Sequence 46, Appl1
8	8	1.6	312	12 US-09-965-422-48	Sequence 48, Appl1
9	8	1.6	316	12 US-09-795-271-56	Sequence 56, Appl1
10	8	1.6	316	12 US-09-912-976-64	Sequence 56, Appl1
11	8	1.6	316	12 US-09-965-422-45	Sequence 45, Appl1
12	8	1.6	666	15 US-10-225-630-10	Sequence 64, Appl1
13	7	1.4	20	14 US-10-032-482-23	Sequence 10, Appl1
14	7	1.4	48	9 US-09-864-761-36243	Sequence 23, Appl1
15	7	1.4	59	9 US-09-764-853-614	Sequence 36243, A

16	7	1.4	59	9 US-09-764-898-271	Sequence 271, App
17	7	1.4	59	15 US-10-073-865-100	Sequence 100, App
18	7	1.4	77	9 US-09-393-634-11	Sequence 11, Appl1
19	7	1.4	77	12 US-10-383-982-11	Sequence 11, Appl1
20	7	1.4	98	15 US-10-144-929-125	Sequence 125, App
21	7	1.4	106	15 US-10-032-482-3	Sequence 3, Appl1
22	7	1.4	135	11 US-09-991-936-38	Sequence 38, Appl1
23	7	1.4	135	11 US-09-991-936-41	Sequence 41, Appl1
24	7	1.4	137	10 US-09-921-397-106	Sequence 106, App
25	7	1.4	144	12 US-10-002-631C-6	Sequence 6, App
26	7	1.4	174	12 US-10-017-161-700	Sequence 700, App
27	7	1.4	181	9 US-09-925-297-778	Sequence 778, App
28	7	1.4	202	10 US-09-738-626-6137	Sequence 6137, Ap
29	7	1.4	209	15 US-10-153-668-424	Sequence 424, App
30	7	1.4	209	15 US-10-153-668-478	Sequence 478, App
31	7	1.4	234	12 US-10-238-075-643	Sequence 643, App
32	7	1.4	253	15 US-10-153-668-314	Sequence 314, App
33	7	1.4	253	15 US-10-153-668-476	Sequence 476, App
34	7	1.4	253	16 US-10-156-136-35	Sequence 35, Appl1
35	7	1.4	282	9 US-09-925-301-1234	Sequence 1234, Ap
36	7	1.4	293	15 US-10-151-336-3	Sequence 3, Appl1
37	7	1.4	306	9 US-09-835-788A-19	Sequence 19, Appl1
38	7	1.4	306	12 US-10-175-042-19	Sequence 19, Appl1
39	7	1.4	321	15 US-10-146-772-12	Sequence 12, Appl1
40	7	1.4	329	15 US-10-208-018-104	Sequence 104, App
41	7	1.4	335	15 US-10-151-336-6	Sequence 6, Appl1
42	7	1.4	336	12 US-10-238-075-305	Sequence 305, App
43	7	1.4	339	9 US-09-729-674-138	Sequence 138, App
44	7	1.4	352	11 US-09-925-299-963	Sequence 963, App
45	7	1.4	352	11 US-09-925-299-963	Sequence 963, App
46	7	1.4	364	9 US-09-815-242-13652	Sequence 13652, A
47	7	1.4	374	15 US-10-255-817-10	Sequence 10, Appl1
48	7	1.4	375	12 US-10-272-983-16	Sequence 16, Appl1
49	7	1.4	375	12 US-10-393-807-16	Sequence 16, Appl1
50	7	1.4	375	15 US-10-318-142-2	Sequence 2, Appl1
51	7	1.4	375	15 US-10-225-567A-494	Sequence 494, App
52	7	1.4	393	12 US-10-091-007A-1234	Sequence 224, App
53	7	1.4	404	16 US-10-080-170-332	Sequence 332, App
54	7	1.4	437	12 US-10-032-588-7130	Sequence 7130, Ap
55	7	1.4	459	15 US-10-156-761-14298	Sequence 14298, A
56	7	1.4	466	12 US-10-238-075-480	Sequence 480, App
57	7	1.4	481	12 US-09-882-227-134	Sequence 134, App
58	7	1.4	487	15 US-10-156-761-14234	Sequence 14234, A
59	7	1.4	487	15 US-10-106-698-4685	Sequence 4685, Ap
60	7	1.4	523	15 US-10-156-761-11594	Sequence 11594, A
61	7	1.4	541	12 US-10-181-157-3	Sequence 3, Appl1
62	7	1.4	542	12 US-10-181-157-1	Sequence 1, Appl1
63	7	1.4	579	15 US-10-011-548-2	Sequence 2, Appl1
64	7	1.4	590	15 US-09-733-183A-4	Sequence 4, Appl1
65	7	1.4	590	12 US-10-342-372-4	Sequence 2, Appl1
66	7	1.4	591	12 US-09-733-183A-2	Sequence 2, Appl1
67	7	1.4	591	12 US-10-342-372-2	Sequence 2, Appl1
68	7	1.4	593	12 US-10-241-872-2	Sequence 17, Appl1
69	7	1.4	604	10 US-09-862-027-17	Sequence 9993, Ap
70	7	1.4	606	15 US-10-156-761-9993	Sequence 6787, Ap
71	7	1.4	625	10 US-09-738-626-6787	Sequence 33, Appl1
72	7	1.4	686	12 US-10-080-608A-33	Sequence 8816, Ap
73	7	1.4	686	12 US-10-370-688-122	Sequence 10084, A
74	7	1.4	760	12 US-10-238-075-943	Sequence 17, Appl1
75	7	1.4	786	11 US-09-950-041-2	Sequence 2, Appl1
76	7	1.4	884	12 US-10-179-766-10	Sequence 10, Appl1
77	7	1.4	1040	10 US-09-929-955-9	Sequence 9, Appl1
78	7	1.4	1050	15 US-10-104-966-9	Sequence 9, Appl1
79	7	1.4	1050	15 US-10-156-761-8816	Sequence 8816, Ap
80	7	1.4	1199	15 US-10-156-761-10084	Sequence 10084, A
81	7	1.4	1884	10 US-09-785-770A-17	Sequence 17, Appl1
82	7	1.4	1907	10 US-09-785-770A-16	Sequence 16, Appl1
83	7	1.4	3011	9 US-09-742-659-4	Sequence 4, Appl1
84	7	1.4	3011	9 US-09-916-359-2	Sequence 2, Appl1
85	7	1.4	3011	10 US-09-238-076-20	Sequence 20, Appl1
86	7	1.4	3011	10 US-09-952-572-9	Sequence 9, Appl1
87	7	1.4	3011	10 US-09-929-955-1	Sequence 1, Appl1
88	7	1.4	3011	10 US-09-747-419-20	Sequence 20, Appl1

```

89 7 1.4 3011 11 US-09-891-894-3 Sequence 3, Appl1
90 7 1.4 3011 11 US-09-995-937-20 Sequence 20, Appl1
91 7 1.4 3011 11 US-09-917-563-20 Sequence 20, Appl1
92 7 1.4 3011 12 US-10-184-150-3 Sequence 3, Appl1
93 7 1.4 3011 14 US-10-104-966-1 Sequence 1, Appl1
94 7 1.4 3011 15 US-10-259-275-20 Sequence 2, Appl1
95 7 1.4 3011 16 US-10-232-643-6 Sequence 6, Appl1
96 7 1.4 3012 10 US-09-238-076-2 Sequence 2, Appl1
97 7 1.4 3012 11 US-09-995-937-2 Sequence 2, Appl1
98 7 1.4 3012 11 US-09-917-563-2 Sequence 109, App
99 7 1.4 3562 12 US-10-341-434-109 Sequence 11, Appl
100 6 1.2 8 10 US-09-910-071-11

```

ALIGNMENTS

```

RESULT 1
US-09-892-851-2
; Sequence 2, Application US/09892851
; Patent No. US20020081682A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-6
; CURRENT APPLICATION NUMBER: US/09/892,851
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/114,060
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: US 60/123,967
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 60/141,271
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-892-851-2

```

```

Query Match 100.0%; Score 515; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKEEKPFGKLSFLMPITHEIKVLPNFMFCITFNTVTRDTKDTIYCAPSSGA 60
DB 1 MTKEEKPFGKLSFLMPITHEIKVLPNFMFCITFNTVTRDTKDTIYCAPSSGA 60
QY 61 EATPFIFEWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPELFFALPTVIYPLRDVL 120
DB 61 EATPFIFEWLVVPCAIIFMLIYAKLSNLSKQALFYAVGPELFFALPTVIYPLRDVL 120
QY 121 HPTFEADRLQALIPPGILGLVALIRNNTFAFYVLAELWGSVMLSMFWGFANETIKHE 180
DB 121 HPTFEADRLQALIPPGILGLVALIRNNTFAFYVLAELWGSVMLSMFWGFANETIKHE 180
QY 121 HPTFEADRLQALIPPGILGLVALIRNNTFAFYVLAELWGSVMLSMFWGFANETIKHE 180
DB 121 HPTFEADRLQALIPPGILGLVALIRNNTFAFYVLAELWGSVMLSMFWGFANETIKHE 180
QY 181 AKRFYALFGIGANTSLASGRAIYWASKLRASVEGVDPWGISIRLLMAMTIYSGVLVMA 240
DB 181 AKRFYALFGIGANTSLASGRAIYWASKLRASVEGVDPWGISIRLLMAMTIYSGVLVMA 240
QY 241 SYMMINKNVLTDPFYAPBEWQKKGAKPMNNKDSFLYDRSPYLLLTLLVIAVIGIC 300
DB 241 SYMMINKNVLTDPFYAPBEWQKKGAKPMNNKDSFLYDRSPYLLLTLLVIAVIGIC 300
QY 301 INLEIETWKSQKLQYFNNDYSEFGNFSFWTGVSVLIMLFYGVNVIRKFGMLTGALV 360
DB 301 INLEIETWKSQKLQYFNNDYSEFGNFSFWTGVSVLIMLFYGVNVIRKFGMLTGALV 360
QY 361 TPVWVLLTGIVFPAIVFRNOASGLVAMFGTTPMLAVVVGAIQNIISKSTKIALPDSIK 420
DB 361 TPVWVLLTGIVFPAIVFRNOASGLVAMFGTTPMLAVVVGAIQNIISKSTKIALPDSIK 420

```

```

QY 421 EMAYIPDQKQKRAIDVVARFCKSGGALIQGLVITCGSIGAMTPYLAIVLFI 480
DB 421 EMAYIPDQKQKRAIDVVARFCKSGGALIQGLVITCGSIGAMTPYLAIVLFI 480
QY 481 AIWLSATKLNKFLAQSALKEQVQAQEDSAPASS 515
DB 481 AIWLSATKLNKFLAQSALKEQVQAQEDSAPASS 515

```

RESULT 2

```

US-10-029-386-28700
; Sequence 28700, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: A60MCA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28700
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR6.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN SWISSPROT HIT: Q9GZK4, EVALUATE 1.00e-36
US-10-029-386-28700

```

```

Query Match 1.6%; Score 8; DB 12; Length 73;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 94 LFYAVGTP 101
DB 42 LFYAVGTP 49

```

RESULT 3

```

US-09-789-697A-21
; Sequence 21, Application US/09789697A
; Patent No. US20020064521A1
; GENERAL INFORMATION:
; APPLICANT: Ellenborn, Joshua D.I.
; APPLICANT: Diamond, Don J.
; TITLE OF INVENTION: p53-Specific T Cell Receptor for Adoptive Immunotherapy
; FILE REFERENCE: 1954-279-11
; CURRENT APPLICATION NUMBER: US/09/789,697A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/183,752
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-697A-21

```

```

Query Match 1.6%; Score 8; DB 9; Length 302;
Best Local Similarity 100.0%; Pred. No. 28;

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLYLMA 240
DB 288 VSGLYLMA 295

RESULT 4
US-09-795-271-57

Sequence 57, Application US/09795271
Publication No. US20030165829A1
GENERAL INFORMATION:
APPLICANT: Padigar, Muralidhara
APPLICANT: Majumder, Kumud
APPLICANT: Burgess, Catherine E
APPLICANT: Vermet, Corine A.M
APPLICANT: Fernandes, Elma
APPLICANT: Shinkets, Richard A
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Casman, Stacie
APPLICANT: Spytek, Kimberly A
APPLICANT: Zerkhusen, Bryan
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-694
CURRENT APPLICATION NUMBER: US/09/795,271
CURRENT FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/185,674
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/185,535
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/186,585
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,604
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,584
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,717
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,716
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,719
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,827
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/218,323
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/218,435
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/220,517
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/223,897
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: 60/260,020
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/264,849
PRIOR FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: 60/186,715
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 57
LENGTH: 303
TYPE: PRT
ORGANISM: Homo sapiens
US-09-795-271-57

Query Match 1.6%; Score 8; DB 12; Length 303;
Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

RESULT 6

DB 265 LFYAVGTP 272

RESULT 5
US-09-965-422-47

Sequence 47, Application US/09965422
Publication No. US20030216545A1
GENERAL INFORMATION:
APPLICANT: Spytek, Kimberly A
APPLICANT: Casman, Stacie
APPLICANT: Padigar, Muralidhara
APPLICANT: Dickson, Kevin
APPLICANT: Vermet, Corine
APPLICANT: Spaderina, Steven K
APPLICANT: Shenoy, Suresh G
APPLICANT: Gerlach, Valerie
APPLICANT: Ellerman, Karen
APPLICANT: Edinger, Shlomit
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
APPLICANT: Li, Li
APPLICANT: Malyankar, Uriel M
APPLICANT: Taylor, Sarah
APPLICANT: Gunther, Erik
APPLICANT: Tchernev, Velizar T
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21401-132
CURRENT APPLICATION NUMBER: US/09/965,422
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/236,286
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/236,284
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 60/237,581
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/238,735
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/240,736
PRIOR FILING DATE: 2000-10-16
PRIOR APPLICATION NUMBER: 60/260,019
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,338
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/262,156
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/262,498
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/263,133
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,691
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/266,109
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/271,634
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 47
LENGTH: 310
TYPE: PRT
ORGANISM: Mus musculus
US-09-965-422-47

Query Match 1.6%; Score 8; DB 12; Length 310;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
DB 274 LFYAVGTP 281

RESULT 6

```
US-10-017-161-508
; Sequence 508, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKUTAMA, YUTAKA
; APPLICANT: ARIYATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-508

Query Match      1.6%; Score 8; DB 12; Length 312;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      94 LFYAVGTP 101
Db      274 LFYAVGTP 281

RESULT 7
US-09-965-422-46
; Sequence 46, Application US/09965422
; Publication No. US20030216545A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Casman, Stacie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Dickson, Kevin
; APPLICANT: Vernet, Corine
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Gerlach, Valerie
; APPLICANT: Ellerman, Karen
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Li, Li
; APPLICANT: Malyanekar, Urial M
; APPLICANT: Taylor, Sarah
; APPLICANT: Gunther, Erik
; APPLICANT: Tchernev, Velizar T
; TITLE OF INVENTION: No. US20030216545A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21401-132
; CURRENT APPLICATION NUMBER: US/09/965,422
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,286
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/236,284
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,581
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,735
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/240,736
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/260,019
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,338
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/262,156
; PRIOR FILING DATE: 2001-01-17

US-09-965-422-48
; Sequence 48, Application US/09965422
; Publication No. US20030216545A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Casman, Stacie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Dickson, Kevin
; APPLICANT: Vernet, Corine
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Gerlach, Valerie
; APPLICANT: Ellerman, Karen
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Li, Li
; APPLICANT: Malyanekar, Urial M
; APPLICANT: Taylor, Sarah
; APPLICANT: Gunther, Erik
; APPLICANT: Tchernev, Velizar T
; TITLE OF INVENTION: No. US20030216545A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21401-132
; CURRENT APPLICATION NUMBER: US/09/965,422
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,286
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/236,284
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,581
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,735
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/240,736
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/260,019
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,338
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/262,156
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/262,498
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/263,133
; PRIOR FILING DATE: 2001-01-19
```

PRIOR APPLICATION NUMBER: 60/263,691
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/266,109
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/271,634
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 312
TYPE: PRT
ORGANISM: Mus musculus
US-09-965-422-48

Query Match 1.6%; Score 8; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 9
US-09-795-271-56
Sequence 56, Application US/09795271
Publication No. US20030165829A1
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Majumder, Kumud
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard A
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Caeman, Stacie
APPLICANT: Spytsek, Kimberly A
APPLICANT: Zethusen, Bryan
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-694
CURRENT APPLICATION NUMBER: US/09/795,271
CURRENT FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/185,674
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/185,535
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: 60/186,585
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,604
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,584
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,717
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,716
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,719
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/186,827
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/218,323
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/218,435
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/220,517
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/223,897
PRIOR FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: 60/260,020
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/264,849
PRIOR FILING DATE: 2001-01-29

PRIOR APPLICATION NUMBER: 60/186,715
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 56
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
US-09-795-271-56

Query Match 1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 10
US-09-912-976-64
Sequence 64, Application US/09912976
Publication No. US20030212255A1
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Mezes, Peter
APPLICANT: Burgess, Catherine
APPLICANT: Caeman, Stacie
APPLICANT: Grose, William M
APPLICANT: Alebrook II, John P
APPLICANT: Lepley, Denise M
APPLICANT: Gerlach, Valerie L
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
APPLICANT: Mishra, Vishnu
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-070
CURRENT APPLICATION NUMBER: US/09/912,976
CURRENT FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/221,336
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/238,333
PRIOR FILING DATE: 2000-10-05
PRIOR APPLICATION NUMBER: 60/260,675
PRIOR FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/271,025
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: 60/278,164
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/280,876
PRIOR FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 99
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 316
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-976-64

Query Match 1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281

RESULT 11
US-09-965-422-45
Sequence 45, Application US/09965422
Publication No. US20030216545A1
GENERAL INFORMATION:

```

; APPLICANT: Spytek, Kimberly A
; APPLICANT: Casman, Stacie
; APPLICANT: Pedigaru, Muralidhara
; APPLICANT: Dickson, Kevin
; APPLICANT: Vernet, Corine
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Gerlach, Valerie
; APPLICANT: Ellerman, Karen
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Li, Li
; APPLICANT: Malpankar, Urial M
; APPLICANT: Taylor, Sarah
; APPLICANT: Gunther, Erik
; APPLICANT: Tchermeyev, Velizar T
; TITLE OF INVENTION: No. US20030216545A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21401-132
; CURRENT APPLICATION NUMBER: US/09/965,422
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,286
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/236,284
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,581
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,735
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/240,736
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/260,019
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,338
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/262,156
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/262,498
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/263,133
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,691
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/266,109
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/271,634
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-965-422-45

Query Match          1.6%; Score 8; DB 12; Length 316;
Best Local Similarity 100.0%; Pred. No. 29;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      94 LFVAVGTP 101
      |||||
Db      274 LFVAVGTP 281

RESULT 12
US-10-225-630-10
; Sequence 10, Application US/10225630
; Publication No. US2003011977A1
; GENERAL INFORMATION:
; APPLICANT: Van Rompaey, Luc
; APPLICANT: Van Es, Helmut H. G.
; APPLICANT: Tonne, Peter H. M.
; APPLICANT: Klaassen, Hubertus J. M.
```

```

; TITLE OF INVENTION: MODULATORS OF BONE HOMEOSTASIS IDENTIFIED IN A HIGH-THROUGHPUT S.C.
; FILE REFERENCE: 25,294-B USA
; CURRENT APPLICATION NUMBER: US/10/225,630
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US 60/314,056
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US 60/356,935
; PRIOR FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 666
; TYPE: PRT
; ORGANISM: human
; US-10-225-630-10

Query Match          1.6%; Score 8; DB 15; Length 666;
Best Local Similarity 100.0%; Pred. No. 59;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      390 GTTPMLA 397
      |||||
Db      364 GTTPMLA 371

RESULT 13
US-10-032-482-23
; Sequence 23, Application US/10032482
; Publication No. US20020197270A1
; GENERAL INFORMATION:
; APPLICANT: Rotter, Irina
; APPLICANT: Cohen, Yuda
; APPLICANT: Wolkowicz, Roland
; APPLICANT: Ruiz, Pedro
; APPLICANT: Erez-Alon, Neta
; APPLICANT: Herxler, Johannes
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR INDUCTION OF ANTI-TUMOR
; FILE REFERENCE: COHEN42
; CURRENT APPLICATION NUMBER: US/10/032,482
; CURRENT FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: US/09/445,602
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: PCT/IL98/00266
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: IL 121041
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-032-482-23

Query Match          1.4%; Score 7; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 22;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      431 QKVKGKA 437
      |||||
Db      14 QKVKGKA 20

RESULT 14
US-09-864-761-36243
; Sequence 36243, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```


GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P209C1
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 100
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-100

Query Match 1.4%; Score 7; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 GLIGLVA 142
Db 7 GLIGLVA 13

RESULT 18
US-09-393-634-11
; Sequence 11, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US2002005197A1el Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(77)
; OTHER INFORMATION: Xaa = any amino acid
US-09-393-634-11

Query Match 1.4%; Score 7; DB 9; Length 77;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LTLTLVI 295
Db 37 LTLTLVI 43

RESULT 19
US-10-383-982-11

; Sequence 11, Application US/10383982
; Publication No. US20030157568A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; TITLE OF INVENTION: SF, a No. US20030157568A1el Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/10/383,982
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US/09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(77)
; OTHER INFORMATION: mouse GR01
; OTHER INFORMATION: Xaa = any amino acid
US-10-383-982-11

Query Match 1.4%; Score 7; DB 12; Length 77;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LTLTLVI 295
Db 37 LTLTLVI 43

RESULT 20
US-10-144-929-125
; Sequence 125, Application US/10144929
; Publication No. US20030069405A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2014P1
; CURRENT APPLICATION NUMBER: US/10/144,929
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US/09/251,329
; PRIOR FILING DATE: 1999-02-17
; PRIOR APPLICATION NUMBER: PCT/US98/17044
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-929-125

Query Match 1.4%; Score 7; DB 15; Length 98;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPLMLAV 398
Db 47 TPLMLAV 53

RESULT 21
US-10-032-482-3

Sequence 3, Application US/10032482
Publication No. US20020197270A1
GENERAL INFORMATION:
APPLICANT: Cohen, Irun
APPLICANT: ROTTER, Varda
APPLICANT: Wolkowicz, Roland
APPLICANT: Ruiz, Pedro
APPLICANT: ERIZ-ALON, Neta
APPLICANT: HEKKEU, Johannes
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR INDUCTION OF ANTI-TUMOR
FILE REFERENCE: COHEN42
CURRENT APPLICATION NUMBER: US/10/032,482
CURRENT FILING DATE: 2002-01-02
PRIOR APPLICATION NUMBER: US/09/445,602
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: PCT/IL98/00266
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: IL 121041
PRIOR FILING DATE: 1997-06-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 106
TYPE: PRT
ORGANISM: Mus musculus
US-10-032-482-3

Query Match 1.4%; Score 7; DB 14; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 QKXGKA 437
|||||
Db 53 QKXGKA 59

RESULT 22
US-09-991-936-38
Sequence 38, Application US/09991936
Publication No. US20030073827A1
GENERAL INFORMATION:
APPLICANT: Brandt, Kevin S.
APPLICANT: Gaines, Patrick J.
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
FILE REFERENCE: FC-6-C1
CURRENT APPLICATION NUMBER: US/09/991,936
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US/09/543,668
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 60/128,704
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 1959
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 135
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-991-936-38

Query Match 1.4%; Score 7; DB 11; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 ELKXLP 29
|||||
Db 64 ELKXLP 70

RESULT 23

US-09-991-936-41
Sequence 41, Application US/09991936
Publication No. US20030073827A1
GENERAL INFORMATION:
APPLICANT: Brandt, Kevin S.
APPLICANT: Gaines, Patrick J.
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
FILE REFERENCE: FC-6-C1
CURRENT APPLICATION NUMBER: US/09/991,936
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US/09/543,668
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 60/128,704
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 1959
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 41
LENGTH: 135
TYPE: PRT
ORGANISM: Ctenocephalides felis
US-09-991-936-41

Query Match 1.4%; Score 7; DB 11; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 ELKXLP 29
|||||
Db 64 ELKXLP 70

RESULT 24
US-09-921-397-106
Sequence 106, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: B4809A, JAZ
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 106
LENGTH: 137
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-921-397-106

Query Match 1.4%; Score 7; DB 10; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 356 TGAIVTP 362
|||||
Db 3 TGAIVTP 9

RESULT 25
US-10-002-631C-6
Sequence 6, Application US/10002631C
Publication No. US20030157486A1
GENERAL INFORMATION:
APPLICANT: Graff, Jonathon M.
APPLICANT: Muenster, Matthew
TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES

FILE REFERENCE: A34943 090495.0243
CURRENT APPLICATION NUMBER: US/10/002.631C
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/300,309
NUMBER OF SEQ ID NOS: 324
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 144
TYPE: PRT
ORGANISM: Homo sapiens
US-10-002-631C-6

Query Match 1.4%; Score 7; DB 12; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 LLLTLVLV 294
Db 64 LLLTLVLV 70

RESULT 26
US-10-017-161-700
Sequence 700, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIYO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017.161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 700
LENGTH: 174
TYPE: PRT
ORGANISM: Homo sapiens
US-10-017-161-700

Query Match 1.4%; Score 7; DB 12; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LLLTLVLV 295
Db 46 LLLTLVLV 52

RESULT 27
US-09-925-297-778
Sequence 778, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: ROSEN ET AL.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925.297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 778
LENGTH: 181
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (145)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (155)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (163)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (169)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-778

Query Match 1.4%; Score 7; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLLASG 200
Db 112 ISLLASG 118

RESULT 28
US-09-738-626-6137
Sequence 6137, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738.626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6137
LENGTH: 202
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6137

Query Match 1.4%; Score 7; DB 10; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GVVSVLI 340
Db 38 GVVSVLI 44

RESULT 29
US-10-153-668-424
Sequence 424, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, GOICHI

APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 424
LENGTH: 209
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-424

Query Match 1.4%; Score 7; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
DB 49 ISLASG 55

RESULT 30
US-10-153-668-478
Sequence 478, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 478
LENGTH: 209
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-478

Query Match 1.4%; Score 7; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
DB 49 ISLASG 55

RESULT 31
US-10-238-075-643
Sequence 643, Application US/10238075
Publication No. US20030146324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolate
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
CURRENT FILING DATE: 2002-09-10
PRIOR APPLICATION NUMBER: 0003145
PRIOR FILING DATE: 2000-03-10
NUMBER OF SEQ ID NOS: 1576
SOFTWARE: PatentIn version 3.1
SEQ ID NO 643
LENGTH: 234
TYPE: PRT
ORGANISM: Escherichia coli
US-10-238-075-643

Query Match 1.4%; Score 7; DB 12; Length 234;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GLVAILR 145
DB 36 GLVAILR 42

RESULT 32
US-10-153-668-314
Sequence 314, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, AKIO
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAMA, Kenya
TITLE OF INVENTION: STATE Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 314
LENGTH: 253
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-314

Query Match 1.4%; Score 7; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200

Db 49 ISLASG 55

```
RESULT 33
US-10-153-668-476
; Sequence 476, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STATE Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153, 668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293, 172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316, 031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328, 403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 476
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-476

Query Match 1.4%; Score 7; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 49 ISLASG 55

RESULT 34
US-10-156-136-35
; Sequence 35, Application US/10156136
; Publication No. US20030128696A1
; GENERAL INFORMATION:
; APPLICANT: NI, et al.
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoding
; RECEPTORS
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; APPLICATION NUMBER: US/10/156,136
; FILING DATE: 29-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,146
```

```
; FILING DATE: <Unknown>
; APPLICATION NUMBER: WO US98/00959
; FILING DATE: 21-JAN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF354PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-10-156-136-35
```

Query Match 1.4%; Score 7; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 49 ISLASG 55

```
RESULT 35
US-09-925-301-1234
; Sequence 1234, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1234
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1234
```

Query Match 1.4%; Score 7; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ISLASG 200
Db 78 ISLASG 84

```
RESULT 36
US-10-151-336-3
; Sequence 3, Application US/10151336
; Publication No. US20030079248A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
```

APPLICANT: Palmer, Kenneth
APPLICANT: Hefferon, Kathleen
APPLICANT: Mor, Tsafir
APPLICANT: Antzen, Charles
TITLE OF INVENTION: Gemini Virus Vectors for Gene Expression in Plants
FILE REFERENCE: 4868/84453
CURRENT APPLICATION NUMBER: US/10/151,336
CURRENT FILING DATE: 2002-05-20
PRIOR APPLICATION NUMBER: US/09/414,276
PRIOR FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 293
TYPE: PRT
ORGANISM: bean yellow dwarf virus
US-10-151-336-3

Query Match 1.4%; Score 7; DB 15; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TPELIF 106
Db 39 TPELIF 45

RESULT 37
US-09-835-788A-19
Sequence 19, Application US/09835788A
Patent No. US2002007458A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: Death Domain-Containing Receptor Polynucleotides, Polypeptides, and
FILE REFERENCE: PTO18P1
CURRENT APPLICATION NUMBER: US/09/835,788A
CURRENT FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: PCT/US00/28666
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/159,585
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/167,246
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 19
LENGTH: 306
TYPE: PRT
ORGANISM: Homo sapiens
US-09-835-788A-19

Query Match 1.4%; Score 7; DB 9; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPMLAV 398
Db 190 TPMLAV 196

RESULT 38
US-10-175-042-19
Sequence 19, Application US/10175042
Publication No. US20030181710A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: Death Domain-Containing Receptor Polynucleotides,
TITLE OF INVENTION: Polypeptides, and
FILE REFERENCE: PTO18P1
CURRENT APPLICATION NUMBER: US/10/175,042
CURRENT FILING DATE: 2002-06-20

PRIOR APPLICATION NUMBER: 09/835,788
PRIOR FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: PCT/US00/28666
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/159,585
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/167,246
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 19
LENGTH: 306
TYPE: PRT
ORGANISM: Homo sapiens
US-10-175-042-19

Query Match 1.4%; Score 7; DB 12; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 TPMLAV 398
Db 190 TPMLAV 196

RESULT 39
US-10-146-772-12
Sequence 12, Application US/10146772
Publication No. US20030124698A1
GENERAL INFORMATION:
APPLICANT: Weiner, David
APPLICANT: Chaplin, Jennifer
APPLICANT: Chi, Ellen
APPLICANT: Milan, Aileen
APPLICANT: Desantis, Grace
APPLICANT: Madden, Mark
APPLICANT: Burk, Mark
TITLE OF INVENTION: Nitrilases
FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
CURRENT APPLICATION NUMBER: US/10/146,772
CURRENT FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/309,006
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US 60/351,336
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/300,189
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 09/751,299
PRIOR FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 60/254,414
PRIOR FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/173,609
PRIOR FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 321
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-12

Query Match 1.4%; Score 7; DB 15; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 KDTLIV 54
Db 7 KDTLIV 13

RESULT 40
US-10-208-018-104
; Sequence 104, Application US/10208018
; Publication No. US20030115632A1
; GENERAL INFORMATION:
; APPLICANT: Lardizabal, Kathryn D
; APPLICANT: Hawkins, Deborah J
; APPLICANT: Thompson, Gregory A
; TITLE OF INVENTION: Diacylglycerol Acyltransferase Proteins
; FILE REFERENCE: 16515.155
; CURRENT APPLICATION NUMBER: US/10/208.018
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: US 10/121,857
; PRIOR FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: US 09/345,461
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,631
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/130,829
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 104
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Glycine max
US-10-208-018-104

Query Match 1.4%; Score 7; DB 15; Length 329;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 194 ISLASG 200
Db 178 ISLASG 184

Search completed: November 25, 2003, 10:13:27
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 06:54:49 ; Search time 53 Seconds

(without alignments)
1542.343 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 515

Sequence: 1 MKTKBKPRGKLRSLFMPH.....AQSALKEGVAGQDSAPASS 515

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched:

1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : A Geneseq 19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	515	100.0	515	21	AAV90265
2	313	60.8	524	20	AAV34951
3	29	5.6	529	20	AAV37153
4	11	2.1	523	21	AAV31017
5	11	2.1	587	21	AAV50990
6	11	2.1	618	21	AAV31016
7	11	2.1	655	21	AAV31015
8	10	1.9	569	21	AAV50991
9	9	1.7	498	21	AAV50992

10	8	1.6	58	22	AAO06244	Human polypeptide
11	8	1.6	137	20	AAV32859	Human heat shock p
12	8	1.6	138	22	ABG13197	Novel human diagno
13	8	1.6	138	22	AAU30541	Novel human secret
14	8	1.6	143	22	ABP76367	Human GENSER prote
15	8	1.6	147	23	ABP41536	Human ovarian anti
16	8	1.6	173	21	AAV51067	Murine TCRbeta pro
17	8	1.6	197	18	AAW20458	H. pylori transmem
18	8	1.6	198	18	AAW20930	H. pylori surface
19	8	1.6	207	23	AAE27063	Mouse TCR Jintubet
20	8	1.6	222	22	AAU33158	Novel human secret
21	8	1.6	247	17	AAE85467	PHLV-L env protein
22	8	1.6	293	7	AAE60238	Sequence of the be
23	8	1.6	293	10	AAE91815	Mammalian T lympho
24	8	1.6	293	18	AAW01532	Cytotoxic T lympho
25	8	1.6	293	21	AAV50114	Murine cytotoxic T
26	8	1.6	302	23	AAE25850	Murine TCR B prote
27	8	1.6	303	20	AAV05403	Killer T-cell rece
28	8	1.6	306	23	ABG73647	Murine beta-TCR mu
29	8	1.6	312	22	AAV71456	Human olfactory re
30	8	1.6	314	20	AAV05728	Mouse A1 T cell re
31	8	1.6	316	22	AAV71665	Human olfactory re
32	8	1.6	316	22	AAV72012	Human olfactory re
33	8	1.6	316	22	AAV72199	Human olfactory re
34	8	1.6	316	22	AAV72973	Olfactory receptor
35	8	1.6	316	23	ABP95755	Human GPCR polypep
36	8	1.6	316	23	ABT04015	Human G-protein co
37	8	1.6	316	23	ABU04016	Human G-protein co
38	8	1.6	317	21	AAV72904	Arabidopsis thalia
39	8	1.6	318	22	AAV72904	Human olfactory re
40	8	1.6	320	22	AAV72622	Murine OR-like pol
41	8	1.6	320	22	AAV72905	Human olfactory re
42	8	1.6	326	22	AAV72626	Murine OR-like pol
43	8	1.6	343	23	AAV71457	Human olfactory re
44	8	1.6	343	22	AAV72371	Human OR-like poly
45	8	1.6	347	22	AAV72625	Murine OR-like pol
46	8	1.6	348	22	AAV72624	Murine OR-like pol
47	8	1.6	351	22	AAV72621	Murine OR-like pol
48	8	1.6	356	20	AAV78160	Human secreted pro
49	8	1.6	356	22	AAV92659	Human protein sequ
50	8	1.6	360	22	AAV78566	Human protein sequ
51	8	1.6	368	21	AAV58321	Arabidopsis thalia
52	8	1.6	378	21	AAV58320	Arabidopsis thalia
53	8	1.6	378	22	ABV11981	Human Slp-2 homolo
54	8	1.6	378	22	AAV79550	Human protein SEQ
55	8	1.6	392	22	ABG12913	Novel human diagno
56	8	1.6	400	10	AAV93189	Novel human diagno
57	8	1.6	400	22	AAU29332	Novel mar regulate
58	8	1.6	430	23	ABG73650	Murine single chai
59	8	1.6	433	22	ABG12912	Novel human diagno
60	8	1.6	443	22	ABG13721	Novel human diagno
61	8	1.6	464	22	AAU28241	Novel human secret
62	8	1.6	467	23	ABP69216	Human polypeptide
63	8	1.6	485	18	AAW23245	Human angiotensino
64	8	1.6	485	18	AAU31120	Novel human secret
65	8	1.6	485	22	AAV67350	Human angiotensino
66	8	1.6	485	22	AAV48944	Human angiotensino
67	8	1.6	485	22	AAV48945	Human angiotensino
68	8	1.6	485	22	AAV48947	Human angiotensino
69	8	1.6	485	22	AAV48948	Human angiotensino
70	8	1.6	485	22	AAV48949	Human angiotensino
71	8	1.6	492	16	AAV85471	Human angiotensino
72	8	1.6	666	21	ABR39453	PTLV-L env glycopr
73	8	1.6	1013	21	AAV84687	Notch-intracellular
74	8	1.6	1013	22	AAE11053	Amino acid sequenc
75	8	1.6	1360	22	ABE61753	Asbha gossypii GT
76	8	1.6	1664	20	AAV95557	Drosophila melanog
77	8	1.6	3067	22	AAU28053	Mus musculus notch
78	8	1.4	16	15	AAV45246	Novel human secret
79	8	1.4	16	15	AAV45247	HCV NS5 region (24
80	8	1.4	16	15	AAV10136	HCV NS5 region (24
81	7	1.4	16	23	ABG79819	T cell epitope/MHC
82	7	1.4	20	20	AAV89167	MHC class I molecu
						Anti-ps3 monoclonal

83 7 1.4 21 23 ABB84104
 84 7 1.4 25 23 AAM51808
 85 7 1.4 30 23 AA084759
 86 7 1.4 37 22 AA009368
 87 7 1.4 48 22 ABG50377
 88 7 1.4 48 22 ABG30346
 89 7 1.4 48 22 ABB35518
 90 7 1.4 48 22 ABB20945
 91 7 1.4 48 22 AAM56335
 92 7 1.4 48 22 AAM68709
 93 7 1.4 48 22 AAM16537
 94 7 1.4 48 22 AAM29017
 95 7 1.4 48 22 AAM04253
 96 7 1.4 48 23 ABB38232
 97 7 1.4 52 21 AAB45031
 98 7 1.4 57 23 ABB03855
 99 7 1.4 59 22 ABB10306
 100 7 1.4 59 22 AAU18124

ALIGNMENTS

Naja naja atra sho
 HCV protease inhib
 HCV Hepcla segment
 Human polypeptide
 Human liver peptid
 Peptide #2997 enco
 Peptide #3024 enco
 Protein #2944 enco
 Human brain expres
 Human bone marrow
 Peptide #2971 enco
 Peptide #3054 enco
 Peptide #2935 enco
 Human peptid enco
 Human secreted pro
 Human CRFX protein
 Human cDNA SEQ ID
 Novel human uterin

RESULT 1

AAV90265 standard; Protein; 515 AA.

22-SEP-2000 (first entry)

C. pneumoniae ATP/ADP translocase protein sequence.

ATP/ADP translocase; Chlamydia infection; diagnosis; therapy.

Chlamydia pneumoniae.

MO200039157-A1.

06-JUL-2000.

22-DEC-1999; 99WO-CA01224.

28-DEC-1998; 98US-0114060.

12-MAR-1999; 99US-0123967.

30-JUN-1999; 99US-0141271.

(CONN-) CONNAUGHT LAB LTD.

Muridin AD, Oomen RP, Wang J, Dunn P;

WPI; 2000-452368/39.

N-PSDB; AAA30922.

Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,

prevention and treatment of Chlamydia infection in mammals -

Claim 16; Fig 1; 81pp; English.

This sequence represents the Chlamydia pneumoniae ATP/ADP translocase

the DNA or protein, are useful for diagnosing, preventing or treating

Chlamydia infection. The sequences can also be used in a method for

the detection of Chlamydia infection. Primers or probes derived from the

DNA sequence are useful in diagnostic tests for detecting Chlamydia

infection.

Sequence 515 AA;

Query Match 100.0%; Score 515; DB 21; Length 515;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTEKPEKGLRSFLMPHTHETLKKVLPMFLMPCITENYTVLRDPTKDTLIVGAPSSGA 60
 DB 1 MTKTEKPEKGLRSFLMPHTHETLKKVLPMFLMPCITENYTVLRDPTKDTLIVGAPSSGA 60
 QY 61 EAIPIFKFVLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 DB 61 EAIPIFKFVLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPFTVIYPLRDVL 120
 QY 121 HPTPEPDRLOALIPGLGLVALLRWTFPAFVYLAELMGSVNLSPMFGFANEITKHE 180
 DB 121 HPTPEPDRLOALIPGLGLVALLRWTFPAFVYLAELMGSVNLSPMFGFANEITKHE 180
 QY 181 AKRFYALFGIGANISILASGRALVWASKLRASVSEGVDPMGISRLMAMTIVSGVLMA 240
 DB 181 AKRFYALFGIGANISILASGRALVWASKLRASVSEGVDPMGISRLMAMTIVSGVLMA 240
 QY 241 SYWMINKNVLTDPRFYNPEEMQKKGAKPKNMKDSFLYLDSPYILLLTLVIAVGIC 300
 DB 241 SYWMINKNVLTDPRFYNPEEMQKKGAKPKNMKDSFLYLDSPYILLLTLVIAVGIC 300
 QY 301 INLEVTWKSQQLQYPMNNDYSEPMGNSEFMTGVSVLIMLFVGGVIRKPGMLGALV 360
 DB 301 INLEVTWKSQQLQYPMNNDYSEPMGNSEFMTGVSVLIMLFVGGVIRKPGMLGALV 360
 QY 361 TPVWVLLTGIVFPALVIFRNOASGLVAMFGTPTPLMLAVVGAIONLSKSTKXALPDSTK 420
 DB 361 TPVWVLLTGIVFPALVIFRNOASGLVAMFGTPTPLMLAVVGAIONLSKSTKXALPDSTK 420
 QY 421 EMAYIPLDQEQKVKRAIDVVARFGKSGGALIQGLLVICSGISGAMPYLAIVLLFI 480
 DB 421 EMAYIPLDQEQKVKRAIDVVARFGKSGGALIQGLLVICSGISGAMPYLAIVLLFI 480
 QY 481 AIWLVGATKLNKFLAQSALKEQVVAQEDSAPASS 515
 DB 481 AIWLVGATKLNKFLAQSALKEQVVAQEDSAPASS 515

RESULT 2

AAV34951 standard; Protein; 524 AA.

13-SEP-1999 (first entry)

Chlamydia pneumoniae transport polypeptide.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

Chlamydia pneumoniae.

WO9927105-A2.

03-JUN-1999.

20-NOV-1998; 98WO-1B01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(BEST) GENSET.

Griffais R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Page 871-872; Disclosure; 1912pp; English.

AAV34584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAX31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX SQ Sequence 524 AA;

Query Match Best Local Similarity 99.6%; Score 313; DB 20; Length 524;

Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPKGKLSFLMPTHTHEIKVLMFLMFCITFNTYVLRDRTKTLIVGAPGSGA 60
 DB 10 MTKTEKPKGKLSFLMPTHTHEIKVLMFLMFCITFNTYVLRDRTKTLIVGAPGSGA 69
 QY 61 EALPFIKFLVNCATIFMLIYAKLSILSKQALFYAVGPPILFFALPFTVYIPRLDV 120
 DB 70 EALPFIKFLVNCATIFMLIYAKLSILSKQALFYAVGPPILFFALPFTVYIPRLDV 129
 QY 121 HPTPEADRLQAIIPGGLGLVAILRMTFAFYVLAELMGSLMLFMGFANEITKHE 180
 DB 130 HPTPEADRLQAIIPGGLGLVAILRMTFAFYVLAELMGSLMLFMGFANEITKHE 189
 QY 181 AKKFPYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMANTIVSGVLMA 240
 DB 190 AKKFPYALFGIGANISLASGRAIVWASKLRASVSEGVDPWGISRLMANTIVSGVLMA 249
 QY 241 SYWINKNVLTDRFVNEEMOKGKAKRMMKOSFLYLDSPYLLTLVIAYGIC 300
 DB 250 SYWINKNVLTDRFVNEEMOKGKAKRMMKOSFLYLDSPYLLTLVIAYGIC 309
 QY 301 INLEVTWKSQQLQYVNNMDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFGMLTGALV 360
 DB 310 INLEVTWKSQQLQYVNNMDYSEFMGNFSFMTGVSVLIMLFVGNVIRKFGMLTGALV 369
 QY 361 TPWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSTK 420
 DB 370 TPWVLLTGIVFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALPDSTK 429
 QY 421 EMAYIPLDQOKYKKAIDVVAARFGKSGALIQGSLVITCGSIGAMTPYLAIVILFIT 480
 DB 430 EMAYIPLDQOKYKKAIDVVAARFGKSGALIQGSLVITCGSIGAMTPYLAIVILFIT 489
 QY 481 AITWLSATKLNKLFIQASALKEQVNAQEDSAPASS 515
 DB 490 AITWLSATKLNKLFIQASALKEQVNAQEDSAPASS 524

RESULT 3

ID AAY37153 standard; Protein: 529 AA.

XX AC AAY37153;

XX DT 07-OCT-1999 (first entry)

XX DE Chlamydia trachomatis transport protein.

XX XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 XX paratrachoma; inclusion conjunctivitis; genital disease; peritrophic;
 XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 XX bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX OS Chlamydia trachomatis.

XX MO9928475-A2.

XX PN

PD 10-JUN-1999.

XX 27-NOV-1998; 98MO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GSET) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 935-936; 1755pp; English.

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC peritrophic, bartholinitis, pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX SQ Sequence 529 AA;

Query Match Best Local Similarity 100.0%; Score 29; DB 20; Length 529;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AYIPLDQOKYKKAIDVVAARFGKSGG 451

DB 422 AYIPLDQOKYKKAIDVVAARFGKSGG 450

RESULT 4

ID AAG13017 standard; Protein: 523 AA.

XX AC AAG13017;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 12353.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.

XX OS Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138549.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151087.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.1%; Score 11; DB 21; Length 523;
Best Local Similarity 100.0%; Pred. No. 0.07; 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDVV 442
DB 414 KVKGAIDVV 424

RESULT 5
ID AAY50990 standard; Protein; 587 AA.

XX AAY50990;

DT 10-MAR-2000 (first entry)

DE A. thaliana ATP1 protein.

KM ADP/ATP translocator; transgenic plant; plasmid; oil; starch;
KW yield increase; ATP1.

OS Arabidopsis thaliana.

PN W0958654-A2.

PD 18-NOV-1999.

PF 12-MAY-1999; 99WO-EP03292.

PR 13-MAY-1998; 98DB-1021442.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Neuhaus E, Moehlmann T, Graeve-Kampfenkel K, Tjaden J, Schell J;
PI Martini N;

DR WPI; 2000-062295/05.

XX Transgenic plants with modified plasmid ADP-ATP translocator activity

PS Example 2; Fig 3; 60pp; German.

CC This invention describes a novel genetically modified transgenic plant
CC cell, where the genetic modification is the introduction of a foreign
CC nucleic acid molecule, which exists, or its expression leads, to increase

CC of plasmid ADP/ATP translocation activity in particular compared to wild
CC type plant cells. Nucleic acid molecules encoding a plasmid ADP/ATP
CC translocator can be used to create transgenic plants and plant
CC cells with increased yields, especially an increased oil and/or starch
CC content. The plants are able to synthesize a modified starch. Plants,
CC such as maize, with an increased oil and/or starch content are of
CC commercial interest especially for the wet milling industry. Increased
CC starch content is important in food crops and in the paper and textile
CC industries. This sequence represents the Arabidopsis thaliana ATP1
CC protein which is used in the method of the invention.

XX Sequence 587 AA;

Query Match 2.1%; Score 11; DB 21; Length 587;
Best Local Similarity 100.0%; Pred. No. 0.078; 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDVV 442
DB 509 KVKGAIDVV 519

RESULT 6
ID AAG13016 standard; Protein; 618 AA.

XX AAG13016;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 12352.

KM Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132486.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0150589.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150864.
PR 26-AUG-1999; 99US-0150866.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155186.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.1%; Score 11; DB 21; Length 618;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDVY 442
Db 509 KVKGAIDVY 519

RESULT 7
AAG13015
ID AAG13015 standard; Protein; 655 AA.

XX AAG13015;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 12351.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123160.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132048.

PR 05-MAY-1999; 99US-0132407.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140354.

PR 28-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145218.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145813.

PR 27-JUL-1999; 99US-0145813.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 28-JUL-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

```

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147415.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149922.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 2.1%; Score 11; DB 21; Length 655;

```

Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 KVKGAIDIV 442
Db 546 KVKGAIDIV 556

RESULT 8
AAV50991
ID AAV50991 standard; Protein; 569 AA.
XX
AC AAV50991;
XX
DT 10-MAR-2000 (first entry)
XX
DE A. thaliana ATP2 protein.
XX
KM ADP/ATP translocator; transgenic plant; plastid; oil; starch;
KM yield increase; ATP2.
XX
OS Arabidopsis thaliana.
XX
XX WO9958654-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99MO-EP03292.
XX
PR 13-MAY-1999; 98DE-1021442.
XX
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Neuhaus E, Moehlmann T, Graeve-Kampfenkel K, Tjaden J, Schnell J,
PI Martini N;
XX
DR WPI; 2000-062295/05.
XX
PT Transgenic plants with modified plastid ADP-ATP translocator activity
PT
XX
XX Example 2; Fig 3; 60pD; German.
XX
CC This invention describes a novel genetically modified transgenic plant
CC cell, where the genetic modification is the introduction of a foreign
CC nucleic acid molecule, which exists, or its expression leads, to increase
CC of plastid ADP/ATP translocation activity in particular compared to wild
CC type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP
CC translocator can be used to create transgenic plants and plant
CC cells with increased yields, especially an increased oil and/or starch
CC content. The plants are able to synthesize a modified starch. Plants,
CC such as maize, with an increased oil and/or starch content are of
CC commercial interest especially for the wet milling industry. Increased
CC starch content is important in food crops and in the paper and textile
CC industries. This sequence represents the Arabidopsis thaliana ATP2
CC protein which is used in the method of the invention.
XX
SQ Sequence 569 AA;

Query Match 1.9%; Score 10; DB 21; Length 569;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKSGLIQQ 456
Db 522 GKSGLIQQ 531

RESULT 9
AAV50992
ID AAV50992 standard; Protein; 498 AA.
XX

```

AC AAY5092; (first entry)
 XX 10-MAR-2000 (first entry)
 DT R. prowazekii ATP homologue TLC protein.
 DE ADP/ATP translocator; transgenic plant; plastid; oil; starch;
 XX yield increase; TLC protein.
 KM Rickettsia prowazekii.
 XX WO958654-A2.
 XX 18-NOV-1999.
 PD 12-MAY-1999; 99WO-EF03292.
 XX 13-MAY-1998; 98DE-1021442.
 XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX Neuhaus E, Moehmann T, Graewe-Kampfenkel K, Tjaden J, Schell J;
 PI Martini N;
 XX WPI; 2000-062295/05.
 DR Transgenic plants with modified plastid ADP-ATP translocator activity
 XX -
 PT Example 2; Fig 3; 60pp; German.
 PS This invention describes a novel genetically modified transgenic plant
 CC cell, where the genetic modification is the introduction of a foreign
 CC nucleic acid molecule, which exists, or its expression leads, to increase
 CC of plastid ADP/ATP translocation activity in particular compared to wild
 CC type plant cells. Nucleic acid molecules encoding a plastid ADP/ATP
 CC translocator can be used to create transgenic plants and plant
 CC cells with increased yields, especially an increased oil and/or starch
 CC content. The plants are able to synthesize a modified starch. Plants,
 CC such as maize, with an increased oil and/or starch content are of
 CC commercial interest especially for the wet milling industry. Increased
 CC starch content is important in food crops and in the paper and textile
 CC industries. This sequence represents the Rickettsia prowazekii TLC
 CC protein which is used in the method of the invention.
 CC
 SQ Sequence 498 AA;
 Query Match 1.7%; Score 9; DB 21; Length 498;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 246 NKNVLTDP 254
 DB 241 NKNVLTDP 249
 RESULT 10
 AAO6244
 ID AAO6244 standard; Protein; 58 AA.
 XX
 AC AAO6244;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human polypeptide SEQ ID NO 20136.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 PA

XX
 PN WO200164835-A2.
 XX 07-SEP-2001.
 PD 26-FEB-2001; 2001WO-US04927.
 XX 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-514938/56.
 DR N-PSDB; AA186175.
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX Claim 20; SEQ ID NO 20136; 1399pp + Sequence Listing; English.
 PS The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 58 AA;
 Query Match 1.6%; Score 8; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 RLQALILPP 135
 DB 25 RLQALILPP 32
 RESULT 11
 AAY32859
 ID AAY32859 standard; Protein; 137 AA.
 XX
 AC AAY32859;
 XX
 DT 01-NOV-1999 (first entry)
 DE Human heat shock protein homologue 2 (HSPH-2) protein sequence.
 XX
 KM Human heat shock protein homologue; HSPH-1; HSPH-2; diagnosis; cancer;
 KM inflammation; AIDS; Crohn's disease; allergy; therapy; drug screening.
 XX
 OS Homo sapiens.
 XX US5945287-A.
 PN 31-AUG-1999.
 PD 19-NOV-1997; 97US-0974546.
 PF 19-NOV-1997; 97US-0974546.
 PR 19-NOV-1997; 97US-0974546.
 XX
 PA (INCY-) INCYTE PHARM INC.

XX
PI Corley NC, Hillman JL, Lal P, Shah P;
XX
DR WPI, 1999-526253/44.
XX N-PSDB; AA211067.
PT New molecular chaperones useful for diagnosis, prevention and
XX treatment of cancer and inflammation
PS Claim 8; Fig 3; 35pp; English.
XX
CC This sequence is the human heat shock protein homologue, HSPH-2 of
CC the invention. The HSPH-1 polynucleotide was isolated from the BRATUT21
CC cDNA library constructed from cancerous brain tissue. The HSPH-2 DNA
CC sequence was isolated from the BRSTNOT18 cDNA library constructed from
CC diseased breast tissue. The HSPH polynucleotides and their complements
CC are useful for diagnosis of conditions or disorders associated with HSPH
CC expression. HSPH complements are also useful as antagonists for
CC prevention or treatment of cancer e.g. adenocarcinoma, breast and skin,
CC and inflammation associated with e.g. AIDS, Crohn's disease and
CC allergies. HSPH-1 and HSPH-2 are useful for drug screening using
CC libraries of compounds. The HSPH-1, HSPH-2 nucleotide sequences, and
CC their complements, are useful for detecting upstream sequences
CC e.g. promoters and regulatory elements. Vectors containing the DNA
CC sequences are useful for the preparation of HSPH proteins and mRNA
CC probes.
XX
SQ Sequence 137 AA;
Query Match 1.6%; Score 8; DB 20; Length 137;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 SLASGRA 202
DB 36 SLASGRA 43
RESULT 12
ABG13197
ID ABG13197 standard; Protein; 138 AA.
XX
XX AC ABG13197;
XX
XX DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #13188.
XX
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI, 2001-639362/73.
XX DR N-PSDB; AA577364.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -

XX
PS Claim 20; SEQ ID NO 43556; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 138 AA;
Query Match 1.6%; Score 8; DB 22; Length 138;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 443 AARFGKSG 450
DB 56 AARFGKSG 63
RESULT 13
AAU30541
ID AAU30541 standard; Protein; 138 AA.
XX
XX AC AAU30541;
XX
XX DT 18-DEC-2001 (first entry)
XX
XX DE Novel human secreted protein #1032.
XX
XX KM Human; vaccination; gene therapy; nutritional supplement;
XX KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX OS Homo sapiens.
XX
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX
XX PF 16-APR-2001; 2001WO-US08656.
XX
XX PR 18-APR-2000; 2000US-0552929.
XX PR 26-JAN-2001; 2001US-0770160.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX DR WPI, 2001-611725/70.
XX
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20; Page 307; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.

XX Sequence 138 AA;
 SQ

Query Match 1.6%; Score 8; DB 22; Length 138;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 SLIAGSRA 202
 DB 33 SLIAGSRA 40
 |||||

RESULT 14
 ABP76367
 ID ABP76367 standard; Protein; 143 AA.
 XX
 AC ABP76367;
 XX
 DT 21-FEB-2003 (first entry)
 XX
 DE Human GENSET protein SEQ ID 917.
 XX
 DE Cytostatic; antiinflammatory; neurotropic; neuroprotective; cardiac;
 KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;
 KW inflammatory disease; immune disorder; neuromuscular; toxicity;
 KW central nervous system; cardiovascular; gastrointestinal.
 OS Homo sapiens.
 XX
 PN WO200283898-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 18-APR-2001; 2001WO-IB00914.
 XX
 PR 18-APR-2001; 2001WO-IB00914.
 XX
 PA (GENSET) GENSET.
 XX
 PI Benjamin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;
 XX
 DR WPI; 2003-075548/07.
 XX
 PT New GENSET polynucleotides and polypeptides, useful for treating heavy
 PT metal toxicity, cancer, inflammatory diseases, immune disorders, and
 PT the neuromuscular, CNS, cardiovascular or gastrointestinal effects of
 PT the toxicity -
 XX
 XX Claim 14; Page 730; 735pp; English.
 XX
 CC The present invention relates to novel GENSET polynucleotides
 CC (AB236404-AB236911) encoding polypeptides (ABP75963-ABP76368). The
 CC polynucleotides and polypeptides are useful in screening and diagnostic
 CC assays for abnormal GENSET expression and/or biological activity. They
 CC are also useful for screening of compounds for treating or preventing
 CC GENSET-related disorders, such as heavy metal toxicity, cancer,
 CC inflammatory diseases, immune disorders, and the neuromuscular, central

CC nervous system (CNS), cardiovascular or gastrointestinal effects of the
 CC toxicity.
 CC
 XX Sequence 143 AA;
 SQ

Query Match 1.6%; Score 8; DB 24; Length 143;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 SLIAGSRA 202
 DB 36 SLIAGSRA 43
 |||||

RESULT 15
 ABP41536
 ID ABP41536 standard; Protein; 147 AA.
 XX
 AC ABP41536;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HSPSF84, SEQ ID NO:2668.
 XX
 DE Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytosstatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CR, Rosen CA;
 XX
 DR WPI; 2002-147878/19.
 XX
 DR N-PSDB; AB054613.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 XX Claim 11; SEQ ID No 2668; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 147 AA;

Query Match 1.6%; Score 8; DB 23; Length 147;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 SLASGRA 202
 DB 46 SLASGRA 53

RESULT 16
 ID AAY51067 standard; protein; 173 AA.
 XX
 AC AAY51067;
 XX
 DT 22-MAR-2000 (first entry)
 XX
 DE Murine TCRbeta protein fragment #2.
 XX
 KM T cell receptor beta; TCRbeta; murine constant region;
 KM immunosuppressor.
 XX
 OS Mus sp.
 XX
 JF11302299-A.
 XX
 PD 02-NOV-1999.
 XX
 PF 21-APR-1998; 98JP-0110607.
 XX
 PR 21-APR-1998; 98JP-0110607.
 XX
 PA (KIRI) KIRIN BREWERY KK.
 XX
 DR WPI; 2000-075345/07.
 XX
 PT T cell receptor beta chain constant region peptide - has
 PT immunosuppressing activity
 XX
 PS Claim 2; Page 10-11; 15pp; Japanese.
 XX
 CC This invention describes a novel polypeptide comprising substantially
 CC part or all of the constant region of T cell receptor beta chain
 CC (TCRbeta) and containing substantially no other regions of TCRbeta and
 CC having immunosuppressing activity. The TCRbeta requires no consideration
 CC of tissue-compatible antigen for the patient to be dosed nor of antigen
 CC specificity. This sequence represents a fragment of the murine TCRbeta
 CC protein described in the method of the invention.

XX
 SQ Sequence 173 AA;

Query Match 1.6%; Score 8; DB 21; Length 173;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGHVTMA 240
 DB 159 VSGHVTMA 166

RESULT 17
 ID AAM20458 standard; Protein; 197 AA.
 XX
 AC AAM20458;
 XX
 DT 11-JUL-1997 (first entry)
 XX
 DE H. pylori transmembrane protein, 36134661.aa.
 XX
 KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KM identification; binding compound; bacterium; life cycle; activator;
 KM bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KM diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96NO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaerd BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67628.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 628; 1481pp; English.
 XX
 CC This sequence represents a H. pylori protein likely to contain three
 CC membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 160 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX
 SQ Sequence 197 AA;

Query Match 1.6%; Score 8; DB 18; Length 197;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 YIHLTTL 293
 DB 178 YIHLTTL 185

RESULT 18
 AAM20930

ID AAW20930 standard; protein; 198 AA.
 AC AAW20930;
 XX
 DT 21-JUL-1997 (first entry)
 DE H. pylori surface or membrane protein, 16ep10117orf7.
 XX
 KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OF, Smith D, Mellgaard BL;
 XX
 DR WPI; 1997-05306/05.
 DR N-PSDB; AAT68183.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 1327; 1481pp; English.
 XX
 CC The present sequence is a H. pylori surface or membrane protein likely
 CC to contain at least three membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 XX
 SQ Sequence 198 AA;
 XX
 QY Query Match 1.6%; Score 8; DB 18; Length 198;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 DB 286 YIIILLTLL 293
 |||||
 179 YIIILLTLL 186
 XX
 RESULT 19
 AAE27063
 ID AAE27063 standard; Protein; 207 AA.
 XX
 AC AAE27063;
 XX
 DT 13-DEC-2002 (first entry)
 DE Mouse TCR Jintubeta2.6-C protein.
 XX

KM Mouse; T cell receptor; TCR; mesenchymal cell growth; cell therapy;
 KM carcinoma; wound healing; intronic J region; joining region; J region;
 KM constant domain; C domain; receptor; Jintubeta2.6-C.
 XX
 OS Mus sp.
 XX
 FH Key
 FT Region 1..19
 FT Region /note= "Intron 5' to Ubeta2.6 region"
 FT Region 20..34
 FT Region /note= "Ubeta2.6 region"
 FT Region 35..38
 FT Region /note= "Cbeta2 region"
 XX
 PN WO20026636-A2.
 PD 29-AUG-2002.
 XX
 PF 20-FEB-2002; 2002WO-IL00130.
 XX
 PR 20-FEB-2001; 2001IL-0241539.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 PA (UYBA-) UNIV BAR-ILAN.
 XX
 PI Zilpori D, Rozenzajn AL, Barda-Saad M, Shav-Tal Y;
 XX
 DR WPI; 2002-667067/71.
 DR N-PSDB; AAD44572.
 XX
 PT New isolated polynucleotide comprising a transcript of a T cell
 PT receptor gene, useful for transfecting mesenchymal cells for wound
 PT healing or suppression of carcinomas -
 XX
 PS Claim 18; Fig 1; 63pp; English.
 XX
 CC The invention relates to polynucleotide transcripts of T cell receptor
 CC (TCR) gene, as well as protein sequences encoded by these transcripts,
 CC and their use in the modulation of mesenchymal cell growth. The
 CC invention also concerns cDNA molecules encoded by a TCR gene lacking V
 CC region sequences and comprising a constant (C) domain and joining (J)
 CC region sequences and a 5' intronic J sequences upstream to the J region
 CC sequence including an in-frame methionine. Polynucleotides of the
 CC invention are useful for transfecting mesenchymal cells for wound
 CC healing or suppression of carcinomas. Antibodies of the invention are
 CC useful as markers of mesenchymal cells. The invention is useful in cell
 CC therapy. The present sequence is mouse TCR Jintubeta2.6-C protein.
 CC
 XX
 SQ Sequence 207 AA;
 XX
 QY Query Match 1.6%; Score 8; DB 23; Length 207;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 DB 233 VSGIVLMA 240
 |||||
 193 VSGIVLMA 200
 XX
 RESULT 20
 AAU33158
 ID AAU33158 standard; Protein; 222 AA.
 XX
 AC AAU33158;
 XX
 DT 18-DEC-2001 (first entry)
 DE Novel human secreted protein #3649.
 XX
 KM Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX

OS	Homo sapiens.
XX	
PN	MO200179449-AA.
PD	25-OCT-2001.
XX	
PF	16-APR-2001; 2001MO-US08656.
XX	
PR	18-APR-2000; 2000US-0552929.
PR	26-JAN-2001; 2001US-0770160.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Dermanac RT;
DR	WPI; 2001-611725/70.
XX	
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic
XX	vaccination, testing and therapy -
XX	
PS	Claim 20; Page 716; 765PP; English.
XX	
CC	The invention relates to novel human secreted polypeptides. The
CC	polypeptides and antibodies to the polypeptides are useful for
CC	determining the presence of or predisposition to a disease associated
CC	with altered levels of polypeptide. The polypeptides are also useful for
CC	identifying agents (agonists and antagonists) that bind to them. Cells
CC	expressing the proteins are useful for identifying a therapeutic agent
CC	for use in treatment of a pathology related to aberrant expression or
CC	physiological interactions of the polypeptide. Vectors comprising
CC	the nucleic acids encoding the polypeptides and cells genetically
CC	engineered to express them are also useful for producing the proteins.
CC	The proteins are useful in genetic vaccination, testing and
CC	therapy, and can be used as nutritional supplements. They may be used to
CC	increase stem cell proliferation; to regulate haematopoiesis; and in
CC	bone, cartilage, tendon and/or nerve tissue growth or regeneration; and
CC	immune suppression and/or stimulation; as anti-inflammatory agents; and
CC	in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC	sequences of novel human secreted proteins of the invention.
XX	
SQ	Sequence 222 AA;
XX	
QY	195 SLASGRA 202
DB	17 SLASGRA 24
XX	
RESULT 21	
XX	AA85467
AC	AA85467 standard; Protein; 247 AA.
XX	
DT	12-MAR-1996 (first entry)
XX	
DE	PHLV-L env protein.
XX	
KW	PTLV-L; HTLV; baboon; vaccine; diagnostic; env protein;
KW	PH969.
XX	
OS	T-lymphotropic primate virus.
XX	
EH	Key
FT	Cleavage-site
FT	Modified-site
FT	Location/Qualifiers
XX	69..70
XX	/note= "env glycoprotein proteolytic cleavage site"
XX	161
XX	/label= N-glycosylation_site
XX	
XX	MO9526405-A1

[illegible]

```

XX  EP180878-A.
PN
XX
PD  14-MAY-1986.
XX
PF  26-OCT-1985; 85SEP-0113658.
XX
PR  31-OCT-1984; 84US-0666988.
PR  13-JUN-1984; 84US-0620122.
PR  01-MAR-1984; 84US-0585333.
PR  22-OCT-1984; 84US-0663809.
PR  27-JUL-1989; 89US-0358997.
XX
PA  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Saito H, Kranz DM, Eisen HN, Tonegawa S;
PI  WPI; 1986-126342/20.
DR  N-PSDB; AAN60167.
XX
PT  New hetero-dimeric t-lymphocyte receptor and its sub-units -
PT  useful in prodn. of T-cell clone specific antibodies for
PT  immunosays, isolation procedures, specific delivery of bound
PT  drugs etc.
XX
PS  Example; Fig 4b; 50pp; English.
XX
CC  T-cell specific cDNA clones were isolated from an alloreactive CTL
CC  clone 2C of Balb/c mouse origin and specific for the D end of the
CC  Balb/C H-2 complex. A library of cDNA was constructed and screened
CC  with hybridisation probes. Two distinct classes of T-cell-specific
CC  cDNA clones whose genes are rearranged in cytotoxic T lymphocytes
CC  (CTLs) were identified (see AAN60166 and AAN60167). A heterodimeric
CC  T lymphocyte receptor comprising an alpha and a beta subunit is
CC  claimed.
CC  (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ  Sequence 293 AA;

Query Match 1.6%; Score 8; DB 7; Length 293;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVVMA 240
DB 279 VSGLVVMA 286

RESULT 23
AAP91815
ID AAP91815 standard; protein; 293 AA.
XX
AC AAP91815;
XX
DT 25-MAR-2003 (updated)
DT 22-MAY-1990 (first entry)
XX
DE Mammalian T lymphocyte receptor beta subunit.
XX
KW T lymphocyte receptor; alpha subunit; cytotoxic T cell; lymphoma;
KW chemotherapy; PHDS11.
XX
OS Mus.
XX
FH Key Location/Qualifiers
FH Domain 12..107
FT /label=variable domain
FT Domain 108..120
FT /label=joining domain
FT Domain 121..266
FT /label=constant domain
FT Domain 267..288

```

```

FT  Domain /label=transmembrane domain
FT 289..293
FT /label=Cytoplasmic domain
FT Disulfide-bond 34..102
FT Disulfide-bond 151..212
FT Disulfide-bond 247..247
FT /note="links alpha chain to beta chain"
XX
PN US874845-A.
XX
PD 17-OCT-1989.
XX
PP 13-JUN-1984; 84US-0620122.
XX
PR 13-JUN-1984; 84US-0620122.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Saito H, Kranz DM, Eisen HN, Tonegawa S;
PI WPI; 1989-377931/51.
DR N-PSDB; AAN93897.
XX
PT Mammalian T lymphocyte receptor sub-unit - with structure derived using
PT cDNA clones derived from functional murine cytotoxic T lymphocyte clone.
XX
PS Disclosure; Page -; 15pp; English.
XX
CC The sequence was deduced from the DNA insert in PHSD11 isolated from the
CC alloreactive CTL clone 2C of BALB.B (mouse) origin.
CC See also AAP91816.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 293 AA;

Query Match 1.6%; Score 8; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVVMA 240
DB 279 VSGLVVMA 286

RESULT 24
AAW01532
ID AAW01532 standard; Protein; 293 AA.
XX
AC AAW01532;
XX
DT 25-MAR-2003 (updated)
DT 14-APR-1997 (first entry)
XX
DE Cytotoxic T lymphocyte clone 2C beta subunit cDNA.
XX
KW Cytotoxic T lymphocyte; CTL; T cell receptor; T helper cell;
KW PHDS11.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FH Peptide 1..11
FT /label=Sig_peptide
FT Protein 12..293
FT /label=Mat_protein
FT Domain 12..107
FT /label= V
FT /note= "variable region (amino acids 1-96 of the
FT 108..120 mature beta-subunit)".
FT /label= J
FT /note= "joining region (amino acids 97-109 of
FT the mature beta-subunit)".

```

```

FT      Region      121..247
FT      /label= C
FT      /note= "constant region (amino acids 110 to approx.
FT      236 of mature beta-subunit)"
FT      Region      266..290
FT      /label= TM
FT      /note= "transmembrane region (amino acids 255-279
FT      of mature beta-subunit)"
FT      Region      289..293
FT      /label= CY
FT      /note= "cytoplasmic region (amino acids 278..284
FT      of mature beta-subunit)"
FT      Disulfide-bond 34..102
FT      Disulfide-bond 151..212
FT      Disulfide-bond 247
FT      /note= "Cys-247 (Cys-236 of mature beta-subunit)
FT      forms a disulphide bond with Cys-234 of
FT      mature gamma subunit"
FT      US5580961-A.
FT      03-DEC-1996.
FT      18-MAR-1994; 94US-0210326.
FT      13-JUN-1984; 84US-0620122.
FT      14-NOV-1988; 88US-0271217.
FT      08-AUG-1991; 91US-0742539.
FT      18-MAR-1994; 94US-0210326.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Eisen HN, Kranz DM, Salto H, Tonegawa S;
FT      WPI, 1997-033602/03.
FT      N-PSDB; AAT46176.
FT      T cell receptor gamma subunit proteins - useful for antibody prodn.
FT      PS      Disclosure; Fig 4A; 19pp; English.
FT      XX      Amino acid sequences for the beta subunit (AAW01532) and gamma
FT      CC      subunit (AAW01533) of a heterodimeric T cell receptor were deduced
FT      CC      from cDNA clones (AAT46176-77) isolated from alloreactive cytotoxic T
FT      CC      lymphocyte (CTL) clone 2C, of BALB.B origin and specific for the D
FT      CC      end of the BALB/c H-2 complex (d haplotype). Genes corresponding
FT      CC      to these cDNAs are expressed and rearranged specifically in T cells
FT      CC      and show homologies to immunoglobulin variable and constant region
FT      CC      genes. The constant region of the beta chain corresponds to the
FT      CC      beta chain of helper T cells and can be used to produce antibodies
FT      CC      useful in the isolation and identification of CTLs and T helper
FT      CC      cells. Recombinant beta and gamma chains can be produced in host
FT      CC      cells and used to raise T cell receptor-specific antibodies useful
FT      CC      e.g. for antibody-bound drugs delivery.
FT      CC      (Updated on 25-MAR-2003 to correct PF field.)
FT      XX      SO      Sequence      293 AA;
FT      Query Match      1.6%; Score 8; DB 18; Length 293;
FT      Best Local Similarity 100.0%; Pred. No. 41;
FT      Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      233 VSGIIVLMA 240
DB      279 VSGIIVLMA 286

```

```

DT      09-FEB-2000 (first entry)
XX      Murine cytotoxic T-lymphocyte (CTL) receptor beta subunit.
XX      CYTOTOXIC; T-lymphocyte; alloreactive; receptor; subunit; beta;
XX      major histocompatibility complex; MHC; antigen; recognition;
XX      immunoglobulin; variable domain; hypervariable region; constant domain;
XX      antibody.
XX      Mus sp.
XX      Key
XX      Peptide      1..11
XX      Protein      /label= signal_peptide
XX      12..293
XX      Domain      /note= "Mature CTL receptor beta subunit"
XX      12..107
XX      Disulfide-bond /note= "variable domain"
XX      34..102
XX      Domain      /note= "Joining region"
XX      108..120
XX      Domain      /note= "Constant region"
XX      121..266
XX      Disulfide-bond /note= "Constant region"
XX      151..212
XX      Disulfide-bond 247
XX      Domain      /note= "Disulphide bond with Cys 252 of gamma subunit"
XX      267..288
XX      Domain      /note= "Transmembrane domain"
XX      289..293
XX      Domain      /note= "Cytoplasmic domain"
XX      US5977321-A.
XX      02-NOV-1999.
XX      08-AUG-1991; 91US-0742539.
XX      13-JUN-1984; 84US-0620120.
XX      14-NOV-1988; 88US-0271217.
XX      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX      PA      Tonge S, Eisen HN, Salto H, Kranz DM;
XX      WPI, 2000-012127/01.
XX      N-PSDB; AA232662.
XX      A new antibody to a T-lymphocyte receptor gamma subunit -
XX      Example; Fig 4A; 14pp; English.
XX      This sequence represents the beta subunit of a murine alloreactive
XX      CC      cytotoxic T-lymphocyte. CTL receptors comprise 2 subunits, beta
XX      CC      and gamma (AAV50115) and are responsible for the recognition
XX      CC      of antigens presented on major histocompatibility complex
XX      CC      (MHC) molecules. The beta and gamma subunits each comprise two
XX      CC      extracellular immunoglobulin (Ig)-like domains (an N-terminal
XX      CC      variable domain and a more C-terminal constant domain, both stabilised
XX      CC      by intramolecular disulphide bonds), a transmembrane domain and a
XX      CC      catenonic cytoplasmic domain. In addition, the two subunits are linked by
XX      CC      an intermolecular disulphide bond. The invention relates to a novel
XX      CC      antibody against the non-hypervariable region of a mammalian CTL
XX      CC      receptor gamma subunit, which is common to CTL receptor subunits of
XX      CC      different specificity. Such antibodies are useful for T-cell receptor
XX      CC      identification and isolation, and for delivering antibody-bound drugs
XX      CC      to a specific cell.
XX      SO      Sequence      293 AA;
XX      Query Match      1.6%; Score 8; DB 21; Length 293;
XX      Best Local Similarity 100.0%; Pred. No. 41;
XX      Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 233 VSGIIVLMA 240
 |||||
 Db 279 VSGIIVLMA 286

RESULT 26
 AAE25850
 ID AAE25850 standard; Protein; 302 AA.

XX AAE25850;

XX 15-NOV-2002 (first entry)

XX Murine TCR B protein.

XX Human leukocyte antigen; HLA; T cell receptor; cytotoxic T lymphocyte;

XX TCR; CTL; adoptive immunotherapy; tumour; gene therapy; cytostatic;

XX murine.

XX Mus musculus.

XX US2002064521-A1.

XX 30-MAY-2002.

XX 22-FEB-2001; 2001US-0789697.

XX 22-FEB-2000; 2000US-183752P.

XX (CITY) CITY OF HOPE.

XX Ellenhorn JDI, Diamond DJ;

XX WPI; 2002-589174/63.

XX N-PSDB; AAD42573.

XX Novel nucleic acid encoding A chain or B chain of hu p53-specific,

XX human leukocyte antigen restricted mu T cell receptor, useful for

XX producing CD8+ cytotoxic lymphocytes for adoptive tumor immunotherapy

XX Claim 1; Page 17-18; 29pp; English.

XX The present invention relates to novel nucleic acids encoding A chain or

XX B chain of human p53-specific, human leukocyte antigen (HLA) restricted

XX murine T cell receptor (TCR). Human CD8+ cytotoxic T lymphocytes (CTLs)

XX transfected with sequences of the invention are useful in adoptive

XX immunotherapy for treating an individual having a tumour that over-

XX expresses p53. Sequences of the invention are useful in gene therapy.

XX The present sequence is murine TCR B protein.

XX Sequence 302 AA;

XX Query Match 1.6%; Score 8; DB 23; Length 302;

XX Best Local Similarity 100.0%; Pred. No. 42;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 VSGIIVLMA 240
 |||||
 Db 288 VSGIIVLMA 295

RESULT 27
 AAY05403
 ID AAY05403 standard; Protein; 303 AA.

XX AAY05403;

XX 02-JUL-1999 (first entry)

XX Killer T-cell receptor protein sequence.

XX Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.

XX Mus musculus.

XX WO9916885-A1.

XX 08-APR-1999.

XX 28-SEP-1998; 98WO-JP04345.

XX 26-SEP-1997; 97JP-0262536.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX (SAIT/) SAITO T.

XX (TAKA/) TAKAHASHI H.

XX Saito T, Takahashi H;

XX WPI; 1999-255096/21.

XX N-PSDB; AAX36391.

XX Killer T-cell receptor peptide specifically recognizing HIV-infected

XX cells

XX Claim 7; Page 60-62; 75pp; Japanese.

XX This sequence is a killer T-cell receptor protein of the invention,

XX which recognises and damages cells infected with human immunodeficiency

XX virus (HIV), especially with HIV-1 IIIB. The receptor can be used as

XX a constituent of compositions for the treatment of HIV infection.

XX Sequence 303 AA;

XX Query Match 1.6%; Score 8; DB 20; Length 303;

XX Best Local Similarity 100.0%; Pred. No. 42;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 VSGIIVLMA 240
 |||||
 Db 289 VSGIIVLMA 296

RESULT 28
 ABG73647

XX ABG73647 standard; Protein; 306 AA.

XX ABG73647;

XX 27-MAR-2003 (first entry)

XX Murine beta-TCR muvbeta-muvbeta.

XX Murine; T cell receptor; TCR; hdm2; T cell response; alpha TCR; beta TCR;

XX antigen-recognising sequence; ARS; fusion construct; cytostatic;

XX apoptotic; tumour; leukaemia; immunisation.

XX Mus musculus.

XX DE10109854-A1.

XX 12-SEP-2002.

XX 01-MAR-2001; 2001DE-1009854.

XX 01-MAR-2001; 2001DE-1009854.

XX (STAN/) STANISLAWSKI T.

XX Theobald M, Voss H, Stanislawski T;

XX WPI; 2002-714556/78.

XX New polypeptide of a murine alpha/beta T-cell receptor, useful for

XX treating tumors and leukemia, induces specific lysis or apoptosis of

PF cells expressing hdm2 protein -
 XX
 PS Claim 1; Page 19-20; 52pp; German.
 XX
 CC This invention describes a novel murine alphabeta T-cell receptor (TCR)
 CC that mediates a hdm2 protein-specific T cell response, a fusion protein
 CC (FP) that includes the TCR and nucleic acid encoding it, alpha or
 CC beta-chains of a TCR that include the antigen-recognition sequence (ARS)
 CC of an antibody specific for aa 81-88 of hdm2 (or its complex with
 CC HLA-A2-specific antibody) and a method for identifying hdm2-specific
 CC antigens. The TCR of the invention has cytostatic and apoptotic activity.
 CC The products of the invention are useful for treatment, prevention and
 CC diagnosis of hdm2-associated diseases, particularly tumours and
 CC leukemia, including use for passive or active immunisation. They can
 CC also be used to screen for therapeutic agents. This sequence represents
 CC the murine beta-TCR muvbeta-mucbeta described in the disclosure of
 CC the invention.
 CC
 SQ Sequence 306 AA;
 Query Match 1.6%; Score 8; DB 23; Length 306;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 VSGELVMA 240
 |||||
 DB 292 VSGELVMA 299
 RESULT 29
 ID AAG71456
 AC AAG71456 standard; Protein; 312 AA.
 XX
 AC AAG71456;
 XX
 DT 30-JUN-2001 (first entry)
 XX
 DE Human olfactory receptor polypeptide, SEQ ID NO: 1137.
 XX
 KW Human, olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation.
 XX
 OS Homo sapiens.
 XX
 PN WO200127158-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 06-OCT-2000; 2000WO-US27582.
 XX
 PR 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 PA (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Bellenson J, Smith D, Lancelot D, Glusman G, Fuchs T, Yanai I;
 XX
 DR WPI; 2001-290713/30.
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 11; Page 658-659; 1857pp; English.
 XX
 CC The present sequence is an olfactory receptor which is encoded by
 CC one of a number of novel polynucleotides. The polynucleotides can be
 CC used in screening for olfactory agonists and antagonists. The methods
 CC allow for the determination of primary scents and the identification
 CC of the odour receptors used to detect these primary scents. The methods
 CC also enable determination of secondary scents and the identification of
 CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.
 CC
 SQ Sequence 312 AA;
 Query Match 1.6%; Score 8; DB 22; Length 312;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 |||||
 DB 274 LFYAVGTP 281
 RESULT 30
 ID AAY05728
 AC AAY05728 standard; Protein; 314 AA.
 XX
 AC AAY05728;
 XX
 DT 19-JUL-1999 (first entry)
 XX
 DE Mouse A1 T cell receptor beta chain.
 XX
 KW T cell receptor beta chain; mouse; transgenic animal;
 KW animal model; immunological tolerance; graft rejection;
 KW tissue grafting; neonatal intolerance; transplantation antigen;
 KW H-Y antigen.
 XX
 OS Mus musculus.
 XX
 PN W09916867-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 30-SEP-1998; 98WO-GB02965.
 XX
 PR 01-OCT-1997; 97GB-0020868.
 XX
 PA (ISIS-) ISIS INNOVATION LTD.
 XX
 PI Cobbold SP, Waldmann H, Zelenika D;
 XX
 DR WPI; 1999-255090/21.
 DR N-PSDB; AAX25358.
 XX
 PT Transgenic non-human mammal having only CD4 positive T cells
 PT specific for at least one transplantation antigen, useful for
 PT studying immunological tolerance
 XX
 PS Disclosure; Page 35-36; 41pp; English.
 XX
 CC The present sequence represents the T cell receptor (TCR) beta
 CC chain from the A1 CD4+ T cell clone isolated from CBA/Ca mice. The
 CC A1 clone recognises the minor histocompatibility antigen H-Y
 CC present in male, but absent in female, mice. cDNA (see AAX25358)
 CC encoding the A1 TCR alpha chain was obtained by RT-PCR (see
 CC also AAX25352-53), and was used in the construction of A1(W)
 CC transgenic mice. The invention relates to a transgenic animal
 CC model comprising TCR alpha and beta chains. It provides a
 CC genetically modified non-human mammal having a population of CD4
 CC positive T cells specific for one or a limited number of selected
 CC antigens, including at least transplantation antigen capable of
 CC rejecting a tissue transplant containing the transplantation
 CC antigen and, if applicable, the other selected antigens. The
 CC animal has TCR genes which encode a TCR specific for the
 CC transplantation antigen. The animal is useful for studying
 CC immunological tolerance, especially the mechanisms of tolerance to,
 CC and the rejection of, tissue grafts, and in pregnancy. The animals

CC are also useful for testing agents for biological activity in
 CC promoting or reducing immunological tolerance.

XX
 CC Sequence 314 AA;

Query Match 1.6%; Score 8; DB 20; Length 314;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 VSGLVVMA 240
 |||||
 Db 300 VSGLVVMA 307

RESULT 31

AAAG71665
 ID AAG71665 standard; Protein; 316 AA.

XX AAG71665;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1346.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

PF 06-OCT-2000; 2000MO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

DR New polynucleotides which encode polypeptides involved in olfactory

XX sensation for identifying olfactory agonists and antagonists -

PT Claim 11, Page 831-832; 1857pp; English.

XX The present sequence is an olfactory receptor which is encoded by

CC one of a number of novel polynucleotides. The polynucleotides can be

CC used in screening for olfactory agonists and antagonists. The methods

CC allow for the determination of primary scents and the identification

CC of the odour receptors used to detect these primary scents. The methods

CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation

CC (also called a scent fingerprint or scent profile), which may be used to

CC re-create and edit scents. Libraries of olfactory receptors are useful

CC for determining the interaction pattern of a composition with the

CC receptors, and can be used for determining differences in the olfactory

CC faculties of different individuals.

Db 274 LFYAVGTP 281

RESULT 32

AAAG72012
 ID AAG72012 standard; Protein; 316 AA.

XX AAG72012;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1693.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

PF 06-OCT-2000; 2000MO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

DR New polynucleotides which encode polypeptides involved in olfactory

XX sensation for identifying olfactory agonists and antagonists -

PT Claim 11, Page 1104-1105; 1857pp; English.

XX The present sequence is an olfactory receptor which is encoded by

CC one of a number of novel polynucleotides. The polynucleotides can be

CC used in screening for olfactory agonists and antagonists. The methods

CC allow for the determination of primary scents and the identification

CC of the odour receptors used to detect these primary scents. The methods

CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such

CC secondary scents. This enables the construction of a scent representation

CC (also called a scent fingerprint or scent profile), which may be used to

CC re-create and edit scents. Libraries of olfactory receptors are useful

CC for determining the interaction pattern of a composition with the

CC receptors, and can be used for determining differences in the olfactory

CC faculties of different individuals.

QY 94 LFYAVGTP 101
 |||||

Db 274 LFYAVGTP 281

RESULT 33

AAAG72199
 ID AAG72199 standard; Protein; 316 AA.

XX AAG72199;

DT 30-JUL-2001 (first entry)

```

DE Human olfactory receptor polypeptide, SEQ ID NO: 1980.
XX
KM Human; olfactory receptor; OR; primary scent determination;
XX secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation.
XX
OS Homo sapiens.
XX
PN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
XX 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENTS.
XX (YEDA) YEDA RES & DEV CO LTD.
XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
DR WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 11; Page 1245-1246; 1857pp; English.
XX
CC The present sequence is an olfactory receptor which is encoded by
CC one of a number of novel polynucleotides. The polynucleotides can be
CC used in screening for olfactory agonists and antagonists. The methods
CC allow for the determination of primary scents and the identification
CC of the odour receptors used to detect these primary scents. The methods
CC also enable determination of secondary scents and the identification of
CC combinations of odour receptors that are involved in detecting such
CC secondary scents. This enables the construction of a scent representation
CC (also called a scent fingerprint or scent profile), which may be used to
CC re-create and edit scents. Libraries of olfactory receptors are useful
CC for determining the interaction pattern of a composition with the
CC receptors, and can be used for determining differences in the olfactory
CC faculties of different individuals.
XX
SQ Sequence 316 AA;
XX
Query Match 1.6%; Score 8; DB 22; Length 316;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY 94 LFYAVGTP 101
|||
|||
|||
Db 274 LFYAVGTP 281
|||
|||
|||
RESULT: 34
ID AAG72973 standard; Protein; 316 AA.
XX
AC AAG72973;
XX
DT 30-JUL-2001 (first entry)
DE Olfactory receptor-like polypeptide, SEQ ID NO: 2655.
XX
KM Human; olfactory receptor; OR; primary scent determination;
XX secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation.
XX
OS Unidentified.
XX
PN WO200127158-A2.
XX
PD 19-APR-2001.

```

Pf	06-OCT-2000; 2000WO-US27582.
Xx	
Pr	08-OCT-1999; 99US-0158615.
Pr	24-FEB-2000; 2000US-0184809.
Xx	
Pa	(DIGI-) DIGISCENTS.
Pa	(YEDA) YEDA RES & DEV CO LTD.
Xx	
Pt	Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
Xx	
Dn	WPI; 2001-290713/30.
Pt	New polynucleotides which encode polypeptides involved in olfactory
Pt	sensation for identifying olfactory agonists and antagonists -
Ps	Example 6; Page 1794-1795; 1857dp; English.
Xx	
Cc	The present sequence is an olfactory receptor polypeptide which was
Cc	used as a query sequence in a database search of olfactory receptor
Cc	(OR)-like sequences. The invention relates to isolated polynucleotides
Cc	encoding polypeptides involved in olfactory sensation. The
Cc	polynucleotides can be used in screening for olfactory agonists and
Cc	antagonists. The methods allow for the determination of primary scents
Cc	and the identification of the odour receptors used to detect these
Cc	primary scents. The methods also enable determination of secondary scents
Cc	and the identification of combinations of odour receptors that are
Cc	involved in detecting such secondary scents. This enables the
Cc	construction of a scent representation (also called a scent fingerprint
Cc	or scent profile), which may be used to re-create and edit scents.
Cc	Literatures of olfactory receptors are useful for determining the
Cc	interaction pattern of a composition with the receptors, and can be
Cc	used for determining differences in the olfactory faculties of different
Cc	individuals.
Sq	Sequence 316 AA;
Oy	Query Match 1.6%; Score 8; DB 22; Length 316; Best Local Similarity 100.0%; Pred. No. 44; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Db	94 LFVAVGTP 101 274 LFVAVGTP 281
Result 35	
Id	ABP95755 standard; Protein: 316 AA.
Xx	ABP95755;
Ac	
Xx	
Dt	06-MAR-2003 (first entry)
Xx	
De	Human GPCR polypeptide SEQ ID NO 320.
Kw	Human; GPCR; G protein coupled receptor; signal transduction; olfactory;
Kw	drug development; gustatory; taste; fragrance; receptor.
Xx	
Os	Homo sapiens.
Xx	
En	WO200216548-A2.
Pd	
Xx	28-FEB-2002.
Pf	30-JUL-2001; 2001MO-IB01446.
Xx	
Pr	04-AUG-2000; 2000JP-0237818.
Pr	13-FEB-2001; 2001JP-0034434.
Xx	
Pa	(NISC-) JAPAN SCI & TECHNOLOGY CORP.
Pt	Haga T, Takeda S, Mitaku S;

XX WPI; 2002-304118/34.
 DR N-PSDB; AB243029.
 XX Database global search for G-protein-coupled receptors, proteins and
 PT encoded genes for studying in vivo signal transduction mechanism and
 PT identifying targets for drug development -
 XX
 PS Claim 10; SEQ ID NO 320; 97pp + Sequence Listing; Japanese.
 XX
 CC The invention relates to a method for screening G-protein-coupled
 CC receptor (GPCR) genes (AB242870-AB243216) and/or GPCR proteins
 CC (AB295596-AB295942) by extracting open-reading frames containing 6-8
 CC transmembrane domains with 250-1000 amino acid residues to give a gene
 CC homologous with a known GPCR gene. The receptor proteins and encoded
 CC genes are useful for studying in vivo signal transduction mechanism and
 CC identifying targets for drug development e.g. based on olfactory and
 CC gustatory receptors in form of agonists and antagonists by screening
 CC intrinsic and extrinsic ligands as bitter taste inhibitors, taste
 CC enhancers and fragrance improvers.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pot_sequences.
 CC
 SQ Sequence 316 AA;
 XX
 XX Query Match 1.6%; Score 8; DB 23; Length 316;
 XX Best Local Similarity 100.0%; Pred. No. 44;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 Db 274 LFYAVGTP 281
 XX
 XX RESULT 36
 XX ID ABJ04015 standard; Protein; 316 AA.
 XX
 AC ABJ04015;
 XX
 DT 25-SEP-2002 (first entry)
 XX
 DE Human G-protein coupled receptor SEQ ID NO: 96.
 XX
 KW Human; G-protein coupled receptor; GPCR; developmental disease;
 KW immune disease; retinal disease; metabolic disorder; feeding disorder;
 KW wasting disorder; infection; cancer; allergy; neurological disorder;
 KW behavioural disorder; reproductive disorder; antidiabetic; cardiant;
 KW antiatherosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN WO200255558-A2.
 XX
 PD 18-JUL-2002.
 XX
 PF 18-DEC-2001; 2001WO-US49112.
 XX
 XX 18-DEC-2000; 2000US-256635P.
 PR 21-DEC-2000; 2000US-257876P.
 PR 04-JAN-2001; 2001US-259743P.
 PR 10-JAN-2001; 2001US-260718P.
 PR 12-JAN-2001; 2001US-261498P.
 PR 08-FEB-2001; 2001US-263689P.
 PR 12-FEB-2001; 2001US-267464P.
 PR 14-MAR-2001; 2001US-275946P.
 PR 23-MAR-2001; 2001US-278150P.
 PR 18-APR-2001; 2001US-284591P.
 PR 23-APR-2001; 2001US-285718P.
 PR 19-JUN-2001; 2001US-299327P.
 PR 16-AUG-2001; 2001US-312902P.

XX (CURA-) CURAGEN CORP.
 PA
 XX Padigaru M, Kekuda R, Li L, Ballinger RA, Casman SJ, Spytek KA;
 PI Colman SD, Vernet CM, Shenoy S, Gusev V, Malyankar UM, Edinger S;
 PI Gerlach V, Smithson G, Stone D, Sciore P, Macdougall JR;
 PI Gunther E, Peyman JA, Ellerman K, Millet I, Tchervet VT;
 PI Anderson DW, Wolenc A;
 XX
 DR WPI; 2002-500840/53.
 DR N-PSDB; ABT04182.
 XX
 CC New G-protein coupled receptor polypeptides, encoding nucleic acids and
 PT immunospecific antibodies, useful for treating cardiomyopathy,
 PT atherosclerosis, diabetes or disorder related to cell signal processing
 PT and metabolic pathway modulation -
 XX
 PS Claim 1; Page 58; 473pp; English.
 XX
 CC The present invention provides the protein and coding sequences of human
 CC G-protein coupled receptors designated GPCR-X. These can be used in the
 CC treatment of developmental diseases, immune diseases, retinal diseases,
 CC metabolic disorders, feeding disorders, wasting disorders, infections,
 CC cancer, allergies, neurological disorders, behavioural disorders and
 CC reproductive disorders. The present sequence is a protein of the
 CC invention.
 CC
 SQ Sequence 316 AA;
 XX
 XX Query Match 1.6%; Score 8; DB 23; Length 316;
 XX Best Local Similarity 100.0%; Pred. No. 44;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 LFYAVGTP 101
 Db 274 LFYAVGTP 281
 XX
 XX RESULT 37
 XX ID ABJ04016 standard; Protein; 316 AA.
 XX
 AC ABJ04016;
 XX
 DT 25-SEP-2002 (first entry)
 XX
 DE Human G-protein coupled receptor SEQ ID NO: 98.
 XX
 KW Human; G-protein coupled receptor; GPCR; developmental disease;
 KW immune disease; retinal disease; metabolic disorder; feeding disorder;
 KW wasting disorder; infection; cancer; allergy; neurological disorder;
 KW behavioural disorder; reproductive disorder; antidiabetic; cardiant;
 KW antiatherosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN WO200255558-A2.
 XX
 PD 18-JUL-2002.
 XX
 PF 18-DEC-2001; 2001WO-US49112.
 XX
 XX 18-DEC-2000; 2000US-256635P.
 PR 21-DEC-2000; 2000US-257876P.
 PR 04-JAN-2001; 2001US-259743P.
 PR 10-JAN-2001; 2001US-260718P.
 PR 12-JAN-2001; 2001US-261498P.
 PR 08-FEB-2001; 2001US-263689P.
 PR 12-FEB-2001; 2001US-267464P.
 PR 14-MAR-2001; 2001US-271021P.
 PR 22-FEB-2001; 2001US-275946P.
 PR 14-MAR-2001; 2001US-278150P.
 PR 23-MAR-2001; 2001US-284591P.
 PR 18-APR-2001; 2001US-285718P.

PR 23-APR-2001; 2001US-285718P.
PR 19-JUN-2001; 2001US-299327P.
PR 16-AUG-2001; 2001US-312902P.
XX
XX (CURA-) CURAGEN CORP.
XX Padigar M, Kekuda R, Li L, Ballinger RA, Spytek KA;
PI Colman SD, Vernet CM, Shenoy S, Gusev V, Malpankar UM, Edlinger S;
PI Gerlach V, Smithson G, Stone D, Sciore P, MacDougall JR,
PI Gunther E, Peyman JA, Ellerman K, Millet I, Tchernev VT;
PI Anderson DW, Wolenc A;
DR WPI; 2002-500840/53.
DR N-PSDB; ABT04183.
XX
XX
XX New G-protein coupled receptor polypeptides, encoding nucleic acids and
PT immunospecific antibodies, useful for treating cardiomyopathy,
PT atherosclerosis, diabetes or disorder related to cell signal processing
PT and metabolic pathway modulation -
XX
XX
XX Claim 1; Page 59; 473pp; English.
XX
XX The present invention provides the protein and coding sequences of human
CC G-protein coupled receptors designated GPCR. These can be used in the
CC treatment of developmental diseases, immune diseases, retinal diseases,
CC metabolic disorders, feeding disorders, wasting disorders, infections,
CC cancer, allergies, neurological disorders, behavioural disorders and
CC reproductive disorders. The present sequence is a protein of the
CC invention.
XX
XX Sequence 316 AA;
SQ
Query March 1.6%; Score 8; DB 23; Length 316;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 LFYAVGTP 101
Db 274 LFYAVGTP 281
RESULT 38
AAG58322
ID AAG58322 standard; Protein; 317 AA.
XX
XX AAG58322;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 75272.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132653.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145293.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150586.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151103.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158359.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161982.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 8; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 VMLSLMFW 169
 Db 282 VMLSLMFW 289

RESULT 39
 AACG72904
 ID AACG72904 standard; Protein; 318 AA.

XX AC AACG72904;
 XX 30-JUL-2001 (first entry)

DE Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2586.

XX KW Human olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation;
 KW human olfactory receptor data exploratorium; HORDE.

XX OS Homo sapiens.

XX PN WO200127158-A2.

XX PD 19-APR-2001.

XX PF 06-OCT-2000; 2000WO-US27562.

XX PR 08-OCT-1999; 99US-0158615.

XX PR 24-FEB-2000; 2000US-0184809.

XX PA (DIGI-) DIGISCENTS.

XX PA (YEDA) YEDA RES & DEV CO LTD.

XX PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX DR WPI; 2001-290713/30.

XX PT New polynucleotides which encode polypeptides involved in olfactory

XX PS sensation for identifying olfactory agonists and antagonists -

XX Example 6; Page 1746-1747; 1857pp; English.

XX The present sequence is a polypeptide from the human olfactory receptor

CC data exploratorium (HORDE). It was used as a query sequence in a database
 CC search of olfactory receptor (OR)-like sequences. The invention relates
 CC to isolated polynucleotides encoding polypeptides involved in olfactory
 CC sensation. The polynucleotides can be used in screening for olfactory
 CC agonists and antagonists. The methods allow for the determination of
 CC primary scents and the identification of the odour receptors used to
 CC detect these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour
 CC receptors that are involved in detecting such secondary scents. This
 CC enables the construction of a scent representation (also called a scent
 CC fingerprint or scent profile), which may be used to re-create and edit
 CC scents. Libraries of olfactory receptors are useful for determining the
 CC interaction pattern of a composition with the receptors, and can be
 CC used for determining differences in the olfactory faculties of different
 CC individuals.

CC Sequence 318 AA;

Query Match 1.6%; Score 8; DB 22; Length 318;

Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Gaps 0;

Matches 8; Conservative 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101

DB 278 LFYAVGTP 285

RESULT 40

AAAG72622 ID AAG72622 standard; Protein; 320 AA.

AC AAG72622;

DT 31-JUL-2001 (first entry)

DE Murine OR-like polypeptide query sequence, SEQ ID NO: 2303.

DE Mouse: olfactory receptor; OR: primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation.

OS Mus musculus.

PN WO200127158-A2.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

DR WPI; 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX Example 6; Page 1553-1554; 1857pp; English.

XX The present sequence is a polypeptide encoded by one of 105 newly mined

XX mouse genes. It was used as a query sequence in a database search of

XX olfactory receptor (OR)-like sequences. The invention relates to isolated

XX polynucleotides encoding polypeptides involved in olfactory sensation.

XX The polynucleotides can be used in screening for olfactory agonists and

XX antagonists. The methods allow for the determination of primary scents

XX and the identification of the odour receptors used to detect these

XX primary scents. The methods also enable determination of secondary scents

XX and the identification of combinations of odour receptors that are

CC involved in detecting such secondary scents. This enables the
 CC construction of a scent representation (also called a scent fingerprint
 CC or scent profile), which may be used to re-create and edit scents.
 CC Libraries of olfactory receptors are useful for determining the
 CC interaction pattern of a composition with the receptors, and can be
 CC used for determining differences in the olfactory faculties of different
 CC individuals.

CC Sequence 320 AA;

Query Match 1.6%; Score 8; DB 22; Length 320;

Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Gaps 0;

Matches 8; Conservative 0; Indels 0; Gaps 0;

QY 94 LFYAVGTP 101

DB 280 LFYAVGTP 287

Search completed: November 25, 2003, 10:05:20
 Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 10:08:07 ; Search time 18 Seconds

(without alignments)
1345,486 Million cell updates/sec

Title: US-09-869-433-2

Perfect score: 2630
Sequence: 1 MTKTEBKPGFLRSFLWPIH.....AQSAIKGEVVAQEDSAPASS 515

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2617	99.5	515	1	TLCL1 CHLBN
2	2147	81.6	528	1	TLCL1 CHLBN
3	2143	81.5	529	1	TLCL1 CHLBN
4	1333	50.7	624	1	TLCL1 ARATH
5	1330.5	50.6	618	1	TLCL2 ARATH
6	1306	49.7	631	1	TLCL1 SOLTR
7	1056	40.2	498	1	TLCA1 RICPR
8	1025.5	39.0	540	1	TLCC1 CHLBN
9	969.5	36.9	501	1	TLCC1 RICPR
10	956.5	36.4	543	1	TLCC2 CHLBN
11	953	36.2	543	1	TLCC2 CHLBN
12	880.5	33.5	512	1	TLCD1 RICPR
13	865	32.9	500	1	TLCE1 RICPR
14	824	31.3	507	1	TLCE1 RICPR
15	135.5	5.2	639	1	LACY1 LEULA
16	133	5.1	669	1	YH06 HAIRN
17	132.5	5.0	463	1	NARK1 ECOLI
18	129	4.9	462	1	NARK1 ECOLI
19	128.5	4.9	528	1	COX1 TRIRU
20	126.5	4.8	512	1	OPUD BACSU
21	126.5	4.8	526	1	COX1 CYACA
22	126	4.8	664	1	PTSA1 STRNU
23	125.5	4.8	641	1	YH06 HAIRN
24	123.5	4.7	394	1	YH06 HAIRN
25	122	4.6	687	1	AFUB1 ACTPL
26	121.5	4.6	894	1	YH06 HAIRN
27	121	4.6	536	1	COX1 ALAMA
28	120	4.6	389	1	YH06 HAIRN
29	119	4.5	389	1	YH06 HAIRN
30	118	4.5	462	1	NARK1 ECOLI
31	117.5	4.5	504	1	YH06 HAIRN
32	116.5	4.4	522	1	COX1 MARPO
33	116	4.4	635	1	NARK1 ECOLI

34	115.5	4.4	397	1	YGED1 ECOLI	P39196 escherichia
35	115	4.4	521	1	COX1 PAPRI	P20374 apis mellif
36	115	4.4	541	1	COX1 PODAN	P20681 podospora a
37	115	4.4	562	1	COX1 THETH	O56408 thermus the
38	115	4.4	873	1	COX1 ACACA	O37370 acanthamoeb
39	114.5	4.4	380	1	CYB PINIM	O81w83 pinguinis i
40	114.5	4.4	593	1	Y917 MYCTU	O05909 mycobacteri
41	114	4.3	557	1	ATKA1 ECOL6	O86fj3 escherichia
42	113.5	4.3	461	1	Y608 HAIRN	O57486 haemophilus
43	112.5	4.3	379	1	CYB BALBN	P41281 balaenopter
44	112.5	4.3	393	1	SRPC1 SYNPN	O50277 synecococc
45	111.5	4.2	380	1	CYB ALALA	O86g23 alle alle (
46	111	4.2	557	1	ATKA1 ECOS7	O8x9f8 escherichia
47	111	4.2	567	1	COX1 EMENT	P00402 emericella
48	110.5	4.2	405	1	YH06 HAIRN	P39352 escherichia
49	110.5	4.2	530	1	COX1 METSE	O35101 metridium s
50	110.5	4.2	557	1	ATKA1 ECOLI	P39359 escherichia
51	110	4.2	447	1	Y418 HAIRN	P46999 haemophilus
52	109.5	4.2	379	1	CYB LAMVI	O36227 lama vicugn
53	109.5	4.2	502	1	NAT21 ECOL6	O86fj3 escherichia
54	109.5	4.2	1107	1	YH06 HAIRN	P39285 escherichia
55	109	4.1	380	1	CYB BALBN	O50277 synecococc
56	108.5	4.1	379	1	CYB BALPH	O79520 haptactes d
57	108.5	4.1	460	1	NU4M BRARE	P24950 balaenopter
58	108.5	4.1	532	1	COX1 CHOCR	O5mi1y brachydanio
59	108	4.1	387	1	CYB FODAN	P48866 chondrus cr
60	108	4.1	433	1	DCUA1 ECOLI	O02655 podospora a
61	107.5	4.1	379	1	CYB MEGNO	P44539 escherichia
62	107.5	4.1	438	1	SHIA1 ECOLI	P41289 megaptera n
63	107	4.1	549	1	YJCG1 ECOLI	P76350 escherichia
64	107	4.1	555	1	CSTA1 AQUAE	P32705 escherichia
65	107	4.1	658	1	S2821 HUMAN	O67304 aquifex aeo
66	106.5	4.0	380	1	CYB ALCTO	O43868 homo sapien
67	106.5	4.0	492	1	YD19 METUA	O8m0a6 alca lorda
68	106.5	4.0	509	1	CYB URILLO	O81w84 uria lomvia
69	106.5	4.0	514	1	YFCC1 HAIRN	O58715 methanococc
70	106.5	4.0	524	1	COX1 BERTU	O05143 protoheca
71	106.5	4.0	524	1	COX1 BERTU	P24794 beta vulgar
72	106.5	4.0	527	1	NU4M ANOGA	O07063 arabidopsis
73	106	4.0	447	1	NU4M ANOGA	P34852 anopheles g
74	105.5	4.0	380	1	CYB GYMTI	P29637 gymnomorpha
75	105.5	4.0	380	1	CYB GYMTI	O81w84 uria lomvia
76	105.5	4.0	383	1	Y9421 SYNVA	P43111 synecocyst
77	105.5	4.0	388	1	SH41 HUMAN	Q13638 homo sapien
78	105.5	4.0	394	1	ARAU1 ECOLI	P33910 escherichia
79	105.5	4.0	524	1	COX1 ORYSA	P45578 oryza sativ
80	105.5	4.0	524	1	COX1 WHEAT	P08741 triticum ae
81	105.5	4.0	528	1	COX1 MAIZE	P08742 zea mays (m
82	105.5	4.0	530	1	COX1 SORBI	P05502 sorghum bic
83	105.5	4.0	622	1	COX1 BACSU	P24010 bacillus su
84	105	4.0	336	1	AGLF1 RHIME	Q92356 rhizobium m
85	105	4.0	416	1	LACY1 KLEPN	P18817 klebsiella
86	105	4.0	468	1	TLCA1 VIRIB	Q9kme2 vibrio chol
87	105	4.0	500	1	YJEM1 ECOLI	O00357 escherichia
88	105	4.0	548	1	TOXA1 COCCA	O00357 escherichia
89	105	4.0	569	1	SPOT1 SPICOL	Q41364 spinacia ol
90	104.5	4.0	342	1	Y7621 METUA	O58172 methanococc
91	104.5	4.0	379	1	CYB LAMGL	Q48891 lama glama
92	104.5	4.0	380	1	CYB BUTBU	O94w88 buteo buteo
93	104.5	4.0	380	1	CYB NUMME	Q35160 numida mele
94	104.5	4.0	461	1	PUCG1 RHOCR	P23462 rhodobacter
95	104.5	4.0	534	1	COX1 RICPR	O54069 rickettsia
96	104.5	4.0	741	1	YBIO1 ECOLI	P75783 escherichia
97	104	4.0	500	1	NU4C1 NEPOL	Q6kxv8 nephrolep
98	104	4.0	587	1	S1321 RAT	P70545 rattus norv
99	104	4.0	1120	1	KEPA1 ECOLI	P73318 escherichia
100	103.5	3.9	368	1	Y8121 AQUAE	O66994 aquifex aeo

RESULT 1

ALIGNMENTS

TLCL1 CHLTPN STANDARD; PRT; 515 AA.
 AC Q928J2; Q9USF8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR ADP1 OR CPN0351 OR CP0408.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 RX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=2030349; PubMed=10871362;
 RA Shira T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CHL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; AE001619; AAD18495.1; -
 DR EMBL; AE002202; AAF38252.1; -
 DR EMBL; AP002546; BAA98559.1; -
 DR PIR; E72089; E72089.
 DR PIR; E86534; E86534.
 DR TIGR; CP0408; -
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KM Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.
 FT CONFLICT 141 141 V -> D (IN REF. 3).
 SQ SEQUENCE 515 AA; 57197 MM; E3F4A52CE008E106 CRC64;
 Query Match 99.5%; Score 2617; DB 1; Length 515;
 Best Local Similarity 99.6%; Pred. No. 5.6e-158;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MTKEEKPFGKLSFELPHTHELEKVLPMFLMFCITFNVTYVDRDRTIYVAPSSGA 60
 DB 1 MTKEEKPFGKLSFELPHTHELEKVLPMFLMFCITFNVTYVDRDRTIYVAPSSGA 60
 QY 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFVAGTPTLFLPPTVPLDVL 120
 DB 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFVAGTPTLFLPPTVPLDVL 120
 QY 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFVAGTPTLFLPPTVPLDVL 120
 DB 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFVAGTPTLFLPPTVPLDVL 120
 QY 121 HPTFADRLQAILPPGLIGVAILRNWTPAFAFYVLAELMGSVMSLMFPGFANEITYIHE 180
 DB 121 HPTFADRLQAILPPGLIGVAILRNWTPAFAFYVLAELMGSVMSLMFPGFANEITYIHE 180
 QY 181 AKRFPYALFGIGANISLASGRATYVASKLRASVEGVDPGICISRLMAMTIVSGLVMA 240
 DB 181 AKRFPYALFGIGANISLASGRATYVASKLRASVEGVDPGICISRLMAMTIVSGLVMA 240
 QY 241 SYWINKNVLTDPFVYVPEEMQKKGAKPRMNMKDSFLYIDRSPYLLTLIVAYGIC 300
 DB 241 SYWINKNVLTDPFVYVPEEMQKKGAKPRMNMKDSFLYIDRSPYLLTLIVAYGIC 300
 QY 301 INLEVTWKSQKLQYNNMDSSEFMGNFSFWTGVSVLIMLVGNGVIRKFGMLTGALV 360
 DB 301 INLEVTWKSQKLQYNNMDSSEFMGNFSFWTGVSVLIMLVGNGVIRKFGMLTGALV 360
 QY 361 TPVWVLLTGIVFPALVIFRQASGLVAMFGTPTPLMVAVMGATONISKSTKVALPSTK 420
 DB 361 TPVWVLLTGIVFPALVIFRQASGLVAMFGTPTPLMVAVMGATONISKSTKVALPSTK 420
 QY 421 EMAYIPDQOKYKGAIVVAAFPKSGGALIQGLVYICSGIGAMTPLYAILLFI 480
 DB 421 EMAYIPDQOKYKGAIVVAAFPKSGGALIQGLVYICSGIGAMTPLYAILLFI 480
 QY 481 AIWIVSATKLNKFLAQSALKEQVAGDSAPASS 515
 DB 481 AIWIVSATKLNKFLAQSALKEQVAGDSAPASS 515
 RESULT 2
 TLCL1 CHLTPN STANDARD; PRT; 528 AA.
 AC Q928J2; Q9USF8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR CT065.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 RX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Ck;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE001281; AAC67656.1; -
 CC PIR; C71561; C71561.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SO SEQUENCE 528 AA; 58117 MW; EAD16A43B95B94B2 CRC64;
 Query Match 81.6%; Score 2147; DB 1; Length 528;
 Best Local Similarity 78.1%; Pred. No. 2.3e-128;
 Matches 410; Conservative 52; Mismatches 49; Indels 14; Gaps 3;

QY 1 MTKEEKPFGKLRSLPFIHTEHKKVLPMLFPCITFNTYVLRDXTDLIVAPSGA 60
 DB 1 MTQAEKPFGLRSLPFIHTEHKKVLPMLFPCISFNTYLRDXTDLIVAPSGA 60
 QY 61 EAIFIKFVLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFFALFPVIVPLRDYL 120
 DB 61 EAIFIKFVLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFFALFPVIVPLRDYL 120
 QY 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 DB 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 QY 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 DB 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 QY 181 AKRFYALFGIGANTSLASGRALYASKLRASVSGVDPWGSLRLMAMTIVGLVMA 240
 DB 181 AKRFYALFGIGANTSLASGRALYASKLRASVSGVDPWGSLRLMAMTIVGLVMA 240
 QY 241 SYMMINKNVLTDPRFYNEEMQKGGAKPKMNKDSFLYDRSPYILLTLVAVGIC 300
 DB 241 SYMMINKNVLTDPRFYNEEMQKGGAKPKMNKDSFLYDRSPYILLTLVAVGIC 300
 QY 241 CYMMNKNRYLTDPRFYNEEL-KAKK-SKPKMSGSESYLLRSPYMLLALLVICYIC 298
 DB 241 CYMMNKNRYLTDPRFYNEEL-KAKK-SKPKMSGSESYLLRSPYMLLALLVICYIC 298
 QY 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFTGVSVYLIMLVGVGNVIRKFGMLTGALV 360
 DB 299 INLEVTWKSQKLQYPMNDYSEFMGNFSFTGVSVYLIMLVGVGNVIRKFGMLTGALV 358
 QY 361 TPWVAVLTGYFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKRYALPSTK 420
 DB 359 TPWVAVLTGYFPALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKRYALPSTK 418
 QY 421 EMAYIPDQOQVKGKALIDVVAARFGSGGALIOGGLVYCGSIGAMTPIAVILPFI 480
 DB 419 EMAYIPDQOQVKGKALIDVVAARFGSGGALIOGGLVYCGSIGAMTPIAVILPFI 478
 QY 481 AIWLVSAIKNKLPLAOSALKEQVAAQ-----EDSAPA 513
 DB 479 MWVLTSAIKNKLPLAOSALKEQVAAQ-----EDSAPA 523

RESULT 3
 TLCL CHLMTU STANDARD; PRT; 529 AA.
 ID TLCL CHLMTU
 AC 09PKX5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP_ATP carrier protein 1 (ADP/ATP translocase 1).
 GN TLCA OR TC0335.
 OS Chlamydia muridarum.
 CC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydia.
 CC NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M09N / Nig9;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brumham R.C., Shen C., Gull S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utecherback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis M09N and Chlamydia
 RL pneumoniae A39.";
 CC Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE002301; AAF39198.1; -
 CC PIR; C81714; C81714.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KM Complete proteome.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 220 240 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 322 342 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 463 483 POTENTIAL.
 SO SEQUENCE 529 AA; 58304 MW; EED13A4C751071C9 CRC64;
 Query Match 81.5%; Score 2143; DB 1; Length 529;
 Best Local Similarity 78.5%; Pred. No. 4.1e-128;
 Matches 408; Conservative 55; Mismatches 49; Indels 8; Gaps 3;

QY 1 MTKEEKPFGKLRSLPFIHTEHKKVLPMLFPCITFNTYVLRDXTDLIVAPSGA 60
 DB 1 MTQAEKPFGLRSLPFIHTEHKKVLPMLFPCISFNTYLRDXTDLIVAPSGA 60
 QY 61 EAIFIKFVLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFFALFPVIVPLRDYL 120
 DB 61 EAIFIKFVLVPCALIFMLIYAKLSNLSKQALFYAVGTEFLFFALFPVIVPLRDYL 120
 QY 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 DB 121 HPTFAADLQALIPSGLGLVAIRNMTFAFYVLAELMGVMSLMFWGFANEITKHE 180
 QY 181 AKRFYALFGIGANTSLASGRALYASKLRASVSGVDPWGSLRLMAMTIVGLVMA 240
 DB 181 AKRFYALFGIGANTSLASGRALYASKLRASVSGVDPWGSLRLMAMTIVGLVMA 240
 QY 241 SYMMINKNVLTDPRFYNEEMQKGGAKPKMNKDSFLYDRSPYILLTLVAVGIC 300
 DB 241 SYMMINKNVLTDPRFYNEEMQKGGAKPKMNKDSFLYDRSPYILLTLVAVGIC 300

DB 241 CYWMNNRYVLDPRFYNPAEL-KAKK-SKPKMKGESPYLLRSPMLLALVLCYVC 298

QY 301 INLEVTWKSQKLOQYPMNNDYSEFMGNFSEWTGVVSLIMLFGVGNIRKFGMLTGALV 360

DB 299 INLEVTWKSQKLOQYPMNNDYSEFMGNFSEWTGVVSLIMLFGVGNIRKFGMLTGALV 358

QY 361 TPVWVLGTGIEFVALVFRNOMSGVAFVGGTTPMLAVVGAIONIKSTFYALFSDTK 420

DB 359 TPVWVLGTGIEFVALVFRNOMSGVAFVGGTTPMLAVVGAIONIKSTFYALFSDTK 418

QY 421 EMAYIPLDQEKVKGAIDVVAARFGKSGALLIOGGLVIGSIGAMTPYLAIVLLEPI 480

DB 419 EMAYIPLDQEKVKGAIDVVAARFGKSGALLIOGGLVIGSIGAMTPYLAIVLLEPI 478

QY 481 AIVLVSAITKMLFLAQSALKRQEV-----AQEASAPAS 514

DB 479 MVWLTSATKMLFLAQSALKRQEV-----AQEASAPAS 518

RESULT 4

TLCL ARATH STANDARD; PRT; 624 AA.

ID TLCL ARATH STANDARD; PRT; 624 AA.

AC Q39002; Q94AN9; Q9C974; 30-MAY-2000 (Rel. 39, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Chloroplast ADP/ATP carrier protein 1, chloroplast precursor (ADP/ATP translocase 1) (adenine nucleotide translocase 1).

GN AATP1 OR AT1G80300 OR F516.5.

OS Arabidopsis thaliana (Mouse-ear cress).

CC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotsids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI TaxID=3702;

OX [1]

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Seedling;

RX MEDLINE=96069943; PubMed=7589569;

RA Kampfenkel K., Mohlmann T., Batz O., van Montagu M., Inze D., Neuhaus H.E.;

RT "Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylate translocator of higher plants.";

RL FEBS Lett. 374:351-355(1995).

RN [2]

RN REVISIONS.

RA Kampfenkel K.;

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RX MEDLINE=21016719; PubMed=1130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y., Buchler E., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldenrich A.D., Haas B., Hansen N.F., Hughes B., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A., Miltchev J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., San H., Salzborg S.L., Schwartz J.R., Shim P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D., Utecherbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";

RL Nature 408:816-820(2000).

RN [4]

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the SSP consortium (Salk/Stanford/PGEC).";

RT Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

RL [5]

RN FUNCTION, AND SUBCELLULAR LOCATION.

RX MEDLINE=9717797; PubMed=9025303;

RA Neuhaus H.E., Thom E., Mohlmann T., Steup M., Kampfenkel K.;

RT "Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L.";

RL Plant J. 11:73-82(1997).

CC -1- FUNCTION: May function as an ATP importer.

CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z49227; CAAB9201.2; -.

DR EMBL; AC018848; AAC52434.1; -.

DR EMBL; AF428316; AA16246.1; -.

DR PIR; A045903; AAK76577.1; -.

DR PIR; E96834; E96834.

DR InterPro; IPR004667; ADP_ATP_car.

DR Pfam; PF03219; TLC; 1.

DR TIGRFPMS; TIGR00769; AAA; 1.

KW Transmembrane; Transport; ATP-binding; Multigene family;

KW Chloroplast; Transit peptide.

FT TRANSIT 1 624

FT CHAIN ? 624

FT TRANSMEM 108 128

FT TRANSMEM 182 202

FT TRANSMEM 240 260

FT TRANSMEM 315 335

FT TRANSMEM 446 466

FT TRANSMEM 545 565

FT CONFLICT 45 94

FT CONFLICT 103 103

FT CONFLICT 197 197

FT CONFLICT 281 281

FT CONFLICT 624 AA; 530302A89AD08745 CRC64;

SEQ

Query Match 50.7%; Score 1333; DB 1; Length 624;

Best Local Similarity 51.8%; Pred. No. 4.9e-77;

Matches 259; Conservative 89; Mismatches 140; Indels 12; Gaps 3;

QY 16 LMPHITHELVKLVDPFIMFCITENYTVLRDTKTLIVGAPGSAEAIPIKFWLVPCA 75

DB 105 IFGEVATLTKKIIPGLGFMFCILFNNTIIRTKOVLVTAQSGSAEIIPIKFWLVPCA 164

QY 76 IIPMLITAKSNITLSKQALPFAVGTPLPILFALPPYIYIPRDVHPREFRDQALPP 135

DB 165 IGFMLLYTKSNVLSKQALPFAVGTPLPILFALPPYIYIPRDVHPREFRDQALPP 224

QY 136 GLGLVILRLRWTPFAAFVLAELGWSVLSLMPGFANETIKIHAKRFVALFGIGANIS 195

DB 225 RPKQPIALIRMSCLTYVAELMGSVVSVLFGFANQITTVDEAKKFYLPFGIGANVA 284

QY 196 LLAAGRAIVASKLRAVSEGVDPWGISLRLMAMTVSGIVLMASTWINKNVLTDPF 255

DB 285 LIFGRITVKYFSNLRKMLPGVDGMVAVSLKAMMSIVGMGLAICLLYWMVARYPLPTRS 344

QY 256 YNPEEMQKKGKAPKPMKMDSPLYIDRSPYIILLTLVLAAYGICVILFTWKSQKLO 315

DB 345 KKKKE-----KPKMGWESLKFVSSPPIRLATLVAAVYISINLVEVTWKSQKLO 396

QY 316 YPNNDYSEFMGNFSEWTGVVSLIMLFGVGNIRKFGMLTGALVTPVWVLGTGIEFVAL 375

```

Db      397 FSPSENEYSAGFWDSTCGVATFTMTL-LSQYFENKYGWGAAKITPTVLLTGVAFPSL 455
Qy      376 VIFPNQASGLVAMGCTPMLAVVGAIONISKSTKALPDSTKMAVYPLDDEQKYG 435
Db      456 ILFGGPPALVAKGMPPLAAVVGALONIFSKAKSLDPDCKEMAYIPLDEDTKYG 515
Qy      436 KAAIDVVAARFGKSGGALIOGGLVTCGSGAMPYLAIVLLFFITAWLSATKLNKFL 495
Db      516 KAAIDVVCNPLGKSGGALIOGPMILSFGSLANSTPYLGMILLVITAMLAASKLEGGF- 574
Qy      496 AQSALKEQEVAAQEDSAPASS 515
Db      575 --NSLRSEELEREMERASS 592

RESULT 5
TLC2_ARATH STANDARD; PRT; 618 AA.
ID      1
AC      P92935; Q9M9E3;
DT      30-MAY-2000 (Rel. 39, Created)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Chloroplast ADP,ATP carrier protein 2, chloroplast precursor (ADP/ATP
translocase 2) (Adenine nucleotide translocase 2).
GN      ATP2 OR AT1G5500 OR T16N11.1.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98206726; PubMed=9546649;
RA      Moehmann T., Thaden U., Schwoeppe C., Winkler H.H., Kampfenkel K.,
RA      Neuhaus H.E.;
RT      "Occurrence of two plastidic ATP/ADP transporters in Arabidopsis
thaliana L. -- molecular characterisation and comparative structural
analysis of similar ATP/ADP translocators from plastids and
Rickettsia prowazekii."
RT      Eur. J. Biochem. 252:353-359 (1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=cv. Columbia;
RA      MEDLINE=21016719; PubMed=11130712;
RA      Theologis A., Ecker J.R., Palm C.J., Federepiel N.A., Kaul S.,
RA      White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA      Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA      Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA      Dunn P., Egu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA      Gill J.E., Goldenrich A.D., Haas B., Hansen N.P., Hughes B., Huzar L.,
RA      Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA      Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA      Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA      Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Matzila A.,
RA      Miltchev J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA      Pei G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA      Sakano H., Salzberrg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA      Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA      Utechtack T., Van Aken S., Vayenberg M., Vyeotskaia V.S., Walker M.,
RA      Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT      "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RT      Nature 406:816-820 (2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      STRAIN=cv. Columbia;
RA      Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
RT      "RIKEN Arabidopsis full length cDNA clones (RATLs) sequenced by the
RT      SSP consortium (Salk/Stanford/IGBC)."
RL      Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.

```

```

RA      Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA      Feldmann K.;
RL      Full-length cDNA from Arabidopsis thaliana."
RT      Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC      -1- CAUTION: Ref.1 sequence differs from that shown due to
CC      framehifts and other sequence problems.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X94626; CAA64329.1; ALT SEQ.
DR      EMBL; AC013453; AAF71976.1; -.
DR      EMBL; AY081350; AAL91239.1; -.
DR      EMBL; AY128844; AAM91244.1; -.
DR      EMBL; AY084374; AAM60955.1; -.
DR      PIR; G86288; G86288.
DR      InterPro; IPR004667; ADP_ATP_car.
DR      Pfam; PF03219; TLC; 1.
DR      TrnRFams; TIGR00769; AAA; 1.
DR      TrnSFams; TIGR00769; AAA; 1.
KW      Transmembrane; Transport; ATP-binding; Multigene family;
KW      Chloroplast; Transl. peptide.
FT      TRANSLAT 1
FT      CHAIN 1
SQ      SEQUENCE 618 AA; 67530 MW; B2D49E77F17A87A CRC64;
Query Match 50.6%; Score 1330.5; DB 1; Length 618;
Best Local Similarity 50.9%; Pred. No. 6.9e-77;
Matches 260; Conservative 92; Mismatches 136; Indels 23; Gaps 5;
Qy      16 LMPHTHELKLVLPFMFPCITPNYVLRDRTKTLIVGARGAGAAIPFKFWLVVCA 75
Db      102 IFGVBVTLTKIIVLGLMFICILENYITLIDTKVAVTAKSSAEIIPFKTWVNLPM 161
Qy      76 IIFMLIYAKLSNLSKQALFYAVGTPLIFPALFPYIYPLRDVLAHTEPADRLQALIP 135
Db      162 IGFMLYTKLSNVLSKQALFYAVGTPLIFPALFPYIYPLRDVLAHTEPADRLQALIP 221
Qy      136 GLLGVALILRNWTFAAVYVLAELMGVMSLMFEPANEITKHEAKRFALFGIGANIS 195
Db      222 RFMGPLAIMIRIWSCLFYVAELMGVMSLMFEPANEITKHEAKRFALFGIGANIS 281
Qy      196 ILASGRAIVASKLRASVSGVDPMGSLRLMAMTIVSGIVMAVYMWINKVLTDPFR 255
Db      282 LIFSGRVKTFSSNRKVLGPGVDGMAVSLKAMSVIGMGALICFLYMWVNR----- 333
Qy      256 YNPEMKGKGGAKPKKAMKDSPLYLDRSPYIILTLVIAVGCINLIETVWSQKIQ 315
Db      334 YVLPPTSKKKKKVAPQGMESLKFIVSSPYIRDLATLVAVGSIIVLEVTWWSKIQ 393
Qy      316 YPNMNDYSEFMNGSFPTGVSVILMFVGGANVIRKGMTLGALVTFVWMLTGIYFPAL 375
Db      394 FSPSENEYSAGFWDSTCGVATFTMTL-LSQYFENKYGWGAAKITPTVLLTGVAFPSL 452
Qy      376 VIFPNQASGLVAMGCTPMLAVVGAIONISKSTKALPDSTKMAVYPLDDEQKYG 435
Db      453 ILFGGPPALVAKGMPPLAAVVGALONIFSKAKSLDPDCKEMAYIPLDEDTKYG 512
Qy      436 KAAIDVVAARFGKSGGALIOGGLVTCGSGAMPYLAIVLLFFITAWLSATKLNKFL 491
Db      513 KAAIDVVCNPLGKSGGALIOGPMILSFGSLANSTPYLGMILLVITAMLAASKLEGGF 572
Qy      492 KFLAQSALKEQF-----VAQEDSAPA 513
Db      573 TLMSERELEREMERASSVKIPVSOED-APS 602

```

RESULT 6
 TLCL SOLTU STANDARD; PRT; 631 AA.
 AC O24381;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Plastidic ATP/ADP-transporter.
 OS Solanum tuberosum (Potato).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 CC NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RA Neubaus E.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST. INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y10821; CAA1785.1; -
 DR PIR: T07420; T07420.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TrEMBL: TIGR00769; AAA; 1.
 DR TrEMBL: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Chloroplast.
 SQ SEQUENCE 631 AA; 68886 MW; 33158380DE2901EA CRC64;
 Query Match 49.7%; Score 1306; DB 1; Length 631;
 Best Local Similarity 52.1%; Pred. No. 2.5e-75;
 Matches 256; Conservative 87; Mismatches 136; Indels 12; Gaps 3;
 Oy 24 LKKVLPMLPFCEITFENYTVLRDTDTLIVGAPSGSGAEIPRIKFWLPPCAIIFMLIYA 83
 Db 111 LKAIIPDLMGFFCIIFNTITLRDTQVLTWTKSSAEIIPFKWVMDPAIGMLIYT 170
 Oy 84 KLSNLSKQALFYAVGTPELIFPFIYVPLRVLPTEPEADRLQALPPLGLGLVAI 143
 Db 171 KLANVLSKQALFYVTLIPFIYVPLRVLPTEPEADRLQALPPLGLGLVAI 230
 Oy 144 LRNMTFAAFYVLAELMGVSLMGMWGRANETTKHEAKRFYALFGIGANISLASGAI 203
 Db 221 LRMSFCIFYVAELMGVSVVSLFWGPNQITTYDEAKRFYPLGLANVALIFSGRTV 290
 Oy 204 VMASLKRASVSGVPMGISLRLLAMNTISGLVVMASVMTNKKVLLDPPRYNDEMQX 263
 Db 291 KYFSSLRSLSLPGVGMALISLKGMMSIVVMGCAICFYWVWNRVVALPTR-----S 342
 Oy 264 GKKGAKPRKMKDSFLYLRSPYILLTLVAVGICINLLEVTWKSQKLQYPMNDYS 323
 Db 343 KKKKVKPRMTMESLKFVSSKYRIDLATLVAVGISINLVETWKSQKLQYPMNDYS 402
 Oy 324 EPMGNFSGTWGVSVLMLFVGNVIRKFGMTGALVTPVWVLTGLIYFPAIVIRNQS 383
 Db 403 SFMGDFSTATGATPTMML-LSQWIFDXYGGAAKIPPTVILLTGVGFFSLIFGABLA 461
 Oy 384 GLVMEFGTPTMLAVAVGAIQNLISKRYALPDSKEMAYPIIDQEKVKGKAIDVYA 443
 Db 462 PTLAKFGTPTMLAAVYVGMQNTFSKAKYSIFDCKEMAYPIIDQEKVKGKAIDVYC 521
 Oy 444 ARFGKSGALLQOGLLVICSGIGANTPYAVLVLLFIILWVLSATKLNKLFASALKEQ 503
 Db 522 NPLKSGGALLQOGLLVICSGIGANTPYAVLVLLFIILWVLSATKLNKLFASALKEQ 578

Oy 504 EVAQEDSAPAS 514
 Db 579 EDLEKEMERAS 589
 RESULT 7
 TLCA RICPR STANDARD; PRT; 498 AA.
 AC P19568;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP ATP carrier protein 1 (ADP/ATP translocase 1).
 CC TLCA OR TLCL OR TLCP OR RP053.
 CC Rickettsia prowazekii.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=90060776; PubMed=2555259;
 RA Williamson L.R., Plano G.V., Winkler H.H., Krause D.C., Wood D.O.;
 RA "Nucleotide sequence of the Rickettsia prowazekii ATP/ADP translocase-
 RT encoding gene.";
 RL Gene 80:269-278(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RA "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [3]
 RP TOPOLOGY.
 RX MEDLINE=91258319; PubMed=1904433;
 RA Plano G.V., Winkler H.H.;
 RA "Identification and initial topological analysis of the Rickettsia
 RT prowazekii ATP/ADP translocase.";
 RL J. Bacteriol. 173:3389-3396(1991).
 RN [4]
 RP TOPOLOGY.
 RX MEDLINE=90334349; PubMed=2165754;
 RA Plano G.V., Wood D.O., Winkler H.H.;
 RA "Rickettsia prowazekii and ATP/ADP translocase. Analysis of gene
 RT fusions encoding beta-galactosidase-ATP/ADP translocase fusion
 RL proteins.";
 RL Ann. N.Y. Acad. Sci. 590:397-407(1990).
 RN [5]
 RP TOPOLOGY.
 RX MEDLINE=99141397; PubMed=9917392;
 RA Alexeyev M.F., Winkler H.H.;
 RA "Membrane topology of the Rickettsia prowazekii ATP/ADP translocase
 RT revealed by novel dual pho-lac reporters.";
 RL J. Mol. Biol. 285:1503-1513(1999).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: M28816; AAA26382.1; -.
DR EMBL: AJ235270; CAA14524.1; -.
DR PIR: J00026; J00026.
DR InterPro: IPR004667; ADP_ATP_car.
DR Pfam: PF03219; TLC; 1.
DR TIGRFAMs: TIGR00769; AAA; 1.
KW Transmembrane; Transport; ATP-binding; Multigene family;
KW Complete proteome.
FT DOMAIN 1 33 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 34 54 PROBABLE.
FT DOMAIN 55 67 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 68 88 PROBABLE.
FT DOMAIN 89 92 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 93 113 PROBABLE.
FT DOMAIN 114 147 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 148 168 PROBABLE.
FT DOMAIN 169 184 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 185 205 PROBABLE.
FT DOMAIN 206 218 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 219 239 PROBABLE.
FT DOMAIN 240 279 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 280 300 PROBABLE.
FT DOMAIN 301 320 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 321 341 PROBABLE.
FT DOMAIN 342 348 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 349 369 PROBABLE.
FT DOMAIN 370 379 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 380 400 PROBABLE.
FT DOMAIN 401 438 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 439 459 PROBABLE.
FT TRANSMEM 460 465 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 466 486 PROBABLE.
FT DOMAIN 487 498 CYTOPLASMIC (PROBABLE).
FT DISULFID 37 85 PROBABLE.
FT NP BIND 436 442 ATP (POTENTIAL).
SQ SEQUENCE 498 AA; 56821 MW; D3C450DD2BC0BE53E CMC64;

Query Match 40.2%; Score 1056; DB 1; Length 498;
Best Local Similarity 41.6%; Pred. No. 1,16-59;
Matches 207; Conservative 106; Mismatches 165; Indels 20; Gaps 7;

QY 2 TKTEKPFKGRSLPHTHEKLVPMFMFCITFNVTVAIDTDTLVAGPSGAE 61
DB 3 TSSENTYLSERKTIWPIBOYENKFLPLAFMFCILINSTLSIDGFWV--TDIGTE 60
QY 62 AIPRIKFWLVPCAIIFMLIYAKLSNLSKQALFYAVGTPFLIFPALFTVIYPLRDVLH 121
DB 61 SISPLKTYIVPSAVIMIIYVKICDILKQENVYVITSFLGFFALFAFLVYPLDLVH 120
QY 122 PTEPADRLQALIPGGLGLVAIIRNWTFAFYVIAELMGVMSLTMFGFANEITKTHEA 181
DB 121 PDHKTISLSLAYNPFKWFIKIVGKMSFASYTIAELMGTMLSLLEWQFANQITKIAEA 180
QY 182 KRFPALFGIGNISILASGRAIWAASKLRASVSGVDPMWISLAL---LMAMTIVGSLV 237
DB 181 KRFPMSFGCLANLALPVTSVVIYGLFHEKTIQI-----VAEHKKEVPLVIMITSSFL 232
QY 238 LMASYMMINKKVLTDPRFYNEEMOKGKGAAPKMNKDSFLYDRSPYILLTLVLVAY 297
DB 233 IILTRYMNKQVLTDPRLHYDALYK--EKTAKALSPFESIKMTFTSKYIVYIALIITAY 290
QY 298 GICINLIEVTWKSQKLQYPMNDYSEBNGFSFWTGVSVLNLVPGVGNIRKFGWLTVG 357
DB 291 GVSVALVEGVKWSKYKELYPTKEAYTIYMGQFOFQGVVAIAFWL--IGSNILRKVSWLTA 349
QY 358 ALVTPVWVLTLGIYEPALVIFRNO--ASGLVAMFGTTPMLAVVGAIONISKSTKYLAF 416
DB 350 AMITPLAMFTLIGAAFPSTIFPDSVIANMLTGLASSPPLTIAVMGIMQNVLSKGVKTSLF 409
QY 417 DSTKEMAYIPLDOCKVYKGAADIVVAARFGKSGGALLIQGLIYICSGIGAM--TPYLAV 474
DB 410 DATGMAYIPLDKOLRVYGAQAAVEVIGGRLGKSGGALLIQSTFFLFPVGFIEATPYFAS 469

```

```

QY 475 ILLEFIAMLVSAFKLNK 492
DB 470 IFFIVILMIFAVKLNK 487

RESULT 8
TLC2_CHLN STANDARD; PRT; 540 AA.
ID TLC2_CHLN
AC Q92700; Q9U010;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADP_ATP carrier protein 2 (ADP/ATP translocase 2).
GN TLCB OR ADT 2 OR CPN0614 OR CP0133.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; Pubmed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Jammal C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; Pubmed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Baas S.,
RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.",
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; Pubmed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.",
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001646; AAD18753.1; -.
DR EMBL: AE002174; AAF38016.1; -.
DR EMBL: AP002547; BAA98821.1; -.
DR PIR: B72056; B72056.
DR PIR: C86567; C86567.
DR TIGR: CP0133; -.
DR InterPro: IPR004667; ADP_ATP_car.
DR Pfam: PF03219; TLC; 1.
DR TIGRFAMs: TIGR00769; AAA; 1.
KW Transmembrane; Transport; ATP-binding; Multigene family;
KW Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.

```

FT TRANSMEM 223 243 POTENTIAL.
 FT TRANSMEM 295 315 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.
 FT TRANSMEM 458 478 POTENTIAL.
 FT TRANSMEM 480 500 POTENTIAL.
 SQ SEQUENCE 540 AA; 60050 MW; 5746023ABE2FABA3 CRC64;

Query Match 39.0%; Score 1025.5; DB 1; Length 540;
 Best Local Similarity 39.4%; Pred. No. 9.8e-58;
 Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

QY 1 MTKNEKPKRSLRSLPIHHEKLVPMFLMFCITENTYVADTDITLIVAGPSGA 60
 DB 1 MOSSVKKPFSRRALCPYKSEFSKVPFLFLAFVGVGCLLNMKDTLVIGSDKA 60
 QY 61 EAPFKEKFWLVPACIIFMLIYAKLSNLSKOALFYAVGTPELFFALPTVIYELRDVL 120
 DB 61 EVIPLKMGIVPGAVITVMYGMGSRYPDTPVYCMMAFLGFFLFAVIIVVGDSTL 120
 QY 121 HPTPEADRLQAILPPGLIGLVAILENMTFAAFYVLAELMGVMSLMWGFANEITKTHE 180
 DB 121 HLNSTADKLOELPQGLRGFTVMRYMSYIYVWSELMSSVLSMLFWGLANOITITTE 180
 QY 181 AKRFPALEGIGANISLASGRAIVASKLK-ASVSEGVDPW-GISRLMAMTIVSGVL 238
 DB 181 AGRFALINTGLNLSICAGEISYMGKOTFYAISFACDSHVSMLNLTMLIT-CSGLIM 239
 QY 239 MASYWINKNV-LTDPFRFVPEEMQKKGK- KPRNMKDSFLYDRSPYI 287
 DB 240 I---WLRRRIHHLITDISIPSRVLAEBGAATANLEKKKPKAKANLFLHIGSRYL 295
 QY 288 LLTLLVLAIGICINLIVTWKSQLQYPMNDYSEPMNGSEFTGVSVLIMLVGCGN 347
 DB 296 LGLIIVSYNVLVHLFEVWVKDQVSOIYSSHVEENGYSRITTLIGVSVLAVALTLGQ 355
 QY 348 VIRRGWLTGALVPMVWMLTGIVEFALVI-FRNOAGSLVMEFTPLMLAVVGAIONI 406
 DB 356 CIRKMGVWGLVPLVPLVWVLSGLFFGTTPAKRDISIFGGVLTGTPMLAMTGKQNV 415
 QY 407 LSKSTKVALPSTKEMAYIPLEDOEKVKGAIDVVAARFKSGGALLIOGGLIVIGSIG 466
 DB 416 LSRGTGFFPOTKEMARIPSPEDKNKGAIDGVSRICKSGSLIYQGLIVIFSSVA 475
 QY 467 AMPTVLAVILFITAIVLSATKLNKFLAQA-----LK-----EOEVAQE 508
 DB 476 ASLVNVALVLLIIMVWIAVAVIGKEYYSRAADAVATLKQPKPESSSIVREAOESVQE 535
 QY 509 DSA 511
 DB 536 EWA 538

RESULT 9
 TLCC_RICPR STANDARD; PRT; 501 AA.
 AC Q9ZD67;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 3 (ADP/ATP translocase 3).
 GN TLCC OR TLCC3 OR RP477.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
 OX NCBI_Taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Magrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.B., Zomorodipour A., Andersson J.O.,
 Sierheltz-Ponten T., Alenmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RL mitochondria".
 CC Nature 336:133-140(1996).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLCC FAMILY.
 CC
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AJ235271; CAA14932.1; -.
 DR PIR; B71707; B71707.
 DR InterPro; IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLCC_1.
 DR TIGRFAMs; TIGR00769; AAA_1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 501 AA; 57182 MW; D1EC540EB6D7E91F CRC64;

Query Match 36.9%; Score 969.5; DB 1; Length 501;
 Best Local Similarity 39.6%; Pred. No. 3e-54;
 Matches 197; Conservative 107; Mismatches 176; Indels 17; Gaps 7;

QY 9 FGKRSFLPIHHEKLVPMFLMFCITENTYVADTDITLIVAGPSGAELIPRIKF 68
 DB 8 FEKKEIIPERKEKELFLPMALMMLCLIFNFAKSIDSLV--PSGAEIISFLKL 65
 QY 69 WLVPACIIFMLIYAKLSNLSKOALFYAVGTPELFFALPTVIYELRDVLHPTPEADR 128
 DB 66 WLVPSCVITITLIVKLSNKLNPYIYISVGTFLFFLFAVIYINODIYHNDAMIN 125
 QY 129 LQAILPPGLIGLVAILENMTFAAFYVLAELMGVMSLMWGFANEITKTHEAKRYALF 188
 DB 126 NLASVFNLMKFIKIGSKMSYALMYISELSAVVIMLMWQFANHIFDPAKARFYPVL 185
 QY 189 GIGANISLASGRAIVASKLKASVSEGV--DPR-----GISRLMAMTIVSGVLMA 240
 DB 186 GWMGNIGLIIAGSVLVFSSGOYIIIDELLTDSYNSSNSNIMLOPIISIIIVAGIIMF 245
 QY 241 SYWINGNVLTDPFRFVPEEMQKKGAKPRNMKDSFLYDRSPYLLTLIVIAVIGIC 300
 DB 246 LFRILNKFIITNS--INVLGVKVAATKTKLALIBESIKLIHSKYIGRIALLIICGLL 303
 QY 301 INLIEVTWKSQKLQYPMNDYSEPMNGSEFTGVSVLIMLVGCGNVIRKFGMLTALV 360
 DB 304 INIVEGPMKAKIKELHENTVDYVFMGMENIMMGISCVTEMI-IGSNILRLGLISALL 362
 QY 361 TPVAVLLTGIVFPALVFRNOAGSLVMEFTPLMLAVVGAIONIISKSTKVALPSTK 420
 DB 363 TPVMLSTGEMFPLIFIEIEIGTCGFENL--LYAIIIGAIIONISKSKSLPSTK 420
 QY 421 EMAYIIPDOEKVKGAIDVVAARFKSGGALLIOGGLIV--CGSIGAMTPVLAVLLF 478
 DB 421 EMAYIIPSLERIKGKAIVAVIGIKRKSIGAFQISLFIITIPATDSIIIVLVFIY 480
 QY 479 IIAIVLSATKLNKFL 495
 DB 481 MNMLMWNIIKLNKEYI 497

RESULT 10
 TLCC2_CHLTR STANDARD; PRT; 540 AA.
 AC O94502;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP,ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCD OR CT495.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_Taxid=813;
 [1]
 RN NCBI_Taxid=813;
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001323; AAC68096.1; -
 CC PIR; E71503; E71503.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRfam; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 361 381 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 FT TRANSMEM 455 475 POTENTIAL.
 FT TRANSMEM 477 497 POTENTIAL.
 SQ SEQUENCE 540 AA; 59764 MW; 126398B58C4A17AB CRC64;
 Query Match 36.4%; Score 956.5; DB 1; Length 540;
 Best Local Similarity 37.2%; Pred. No. 2.2e-53;
 Matches 197; Conservative 116; Mismatches 193; Indels 23; Gaps 7;
 QY 4 TEEKPFGRLRSFIMPIHHEHLKVLPMFLMPCITPNTVLRDPTDILVAGAPSGAAI 63
 DB 3 SEVSFSFSGYFPPIYSEBSKPLPLFLAFVGVNALLKTKYDLSLVASRGAGVI 62
 QY 64 PFIKFLVPCAIIFMLIYAKISLNSKQALFYAAGTFFLLFPAFPVPIYPLRDVHPT 123
 DB 63 PFLKVGIVPGAVIVMTIYGMRSRYSGIVFISLVGFLGFPALFAVIVPIGDAHLN 122
 QY 124 EPADRLQAILPGLIGLVAIIIRNMTFAAFYVLAELWGSVMSLMEWGFANETIKHEKR 183
 DB 123 KLAAKLOSILPPGGGCFVMTQWYSYSLYVMSISIVSTLFWGVANHITSVREGR 182
 QY 184 FYALFGGANISLASGRAIYWSKLRASVSEGVDPG-GISRLMAMTIVSGIYMAS 241
 DB 183 FYALINTGLNASSVPAGEVSIWLGPNVIAFPMAVDPHVLLNTLLIVLAGVTL-- 239
 QY 242 YMWINKNVLTDPFYNP-----EMQKKGAKPKMMKDSFLYIDRSPYLLTL 292
 DB 240 YLYQGLDRIMDETSMLEKGLAEMSVAQLEKKKKRSKAK--AKSLFALLLSRYLLGIV 297

QY 293 LVIAVICINLIEYTWKSQLEQYPMNDYSEFMGNFSFWTGVSVLIMLFGVGNVIRKF 352
 DB 298 VVLSYLVTHLFEVWMDQCVRIYASRVEFSYMSRTTLTGIVSALGIAAGQTLNRW 357
 QY 353 GMLTGAIVTPVWVLLTGIVFE-ALVIFPNQASGLVAMFGTTPMLAAVGAIONLSKST 411
 DB 358 GWITGALVPLTMTLITGALFFGAIYAVKGDMMITGSLIGISPLVTMLMGVQVFSRAI 417
 QY 412 KYALPDSIKEMAYIPLDQEKVKGAIDVVAAPFGSGGALLIQGLIVIGSIGAMTPY 471
 DB 418 KFTYFDQTKEMAFIPLEDEKNGVKAIDGVISRVKSGSGSLVYQGLIIFSSVAASLNA 477
 QY 472 LAVLLPFIATWVSATKNTLFLAQA-----LKGQVAGQESAPAS 514
 DB 478 ITIVLLALGSIWIVIAMLGRYAKTAEALFRVNVSESDVLEERBAS 526
 RESULT 11
 ID TLC2_CHLMU STANDARD, PRT, 543 AA.
 AC Q9PJUP6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE ADP,ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCD OR TC0782.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_Taxid=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / Nig9;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwynn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE002346; AAF39585.1; -
 CC PIR; F81665; F81665.
 DR TIGR; TC0782; -
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRfam; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 361 381 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 477 497 POTENTIAL.

SQ SEQUENCE 543 AA; 59672 MW; 6291A671F1924D7D CRC64;
 Query Match 36.2%; Score 953; DB 1; Length 543;
 Best Local Similarity 37.5%; Pred. No. 3.6e-53;
 Matches 199; Conservative 11; Mismatches 197; Indels 24; Gaps 7.

4 TEKKFGKLRSLMFIHTELKVLPMFLMFCITFNVTVDLDTDTLIVGAPGSGAIAI 63
 3 SEVTFKFSRRKRRFFPHKSEFPKIFPLLIAPFVGPNVSLKTTDSDVLASRGAAVI 62
 64 PEIKFVLVPCAIIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPTVYPLRDVLHT 123
 63 PFLKVMGIVGAVITIMTIYGMMSCRSGFVPCALVGGFLSPFALFACIYPMGALHIN 122
 124 EPADRLQALIPPGILGLVALIRNTPFAFYVLAELMGSMLSLWNGFANETTKHEAKR 183
 123 GLAALQTLIPRGAGFVVMQVMSYSLYVMSLMSVSLTFLWGLANHTSVREAGR 182
 184 FYALFGIGANISILASGRAIWAASK-LRASVSEGVDPV-GISRLMLMTIVSGLVMAAS 241
 183 FYALINVGAINVSIYAGETISLWGHKHTLIPSSMAVDANHGVLNITLIIYAAGLIL- 239
 242 YWINKNVLTDP-----RFYNPEEMQKGGKAKPRMNMKOSFLYIDRSPLYLLTL 292
 240 YLYRLDHLTEAPVLGDGVSEMSVAOLKQEKK--RPKAKXSLSLVLPFSRYLMGIAY 297
 293 LVIANGICINLIEVTKWSQKLQYPMNDYSEFMGNSEFNGVSVLIMLFGVGNVIRKF 352
 298 VLAALNLAHLELVVMKQVCOIYSSRVEFNSYMRITAFGTISALAGVFAAGOSIRRW 357
 353 GMLTGLATVPWVLLTGLIVF-ALVIFRNQASGLVAMFGTPEMLAVVGAIONILSKST 411
 358 GNTVALITLPTLMLITGGLFAGIYAVKGDAMILGFLGFSVLITANLGGVQNVFSAL 417
 412 KYALFDSFKEMAYIPLEDOQVKGKAAIDVVAARFGSGGALLIOGLIVIGSIGAMTPY 471
 418 KFYFDQTKEMAFIPLEDEKQYKAAIDGVISRVKSGSLVYQALLIFSSVADCNNA 477
 472 LAVILFLIATMVSAATKNTKFLAQS-----ALKQEVAAQEDSARASS 515
 478 ITVILLALGGMIVWVWMLGKEYSVRTALGKARAEPSPLODEDESRSS 528

RESULT 12
 TLCD_RICPR STANDARD; PRT; 512 AA.
 AC Q92D47;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 4 (ADP/ATP translocase 4).
 GN TLCD OR TLCD4 OR RP500.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sicheitiz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RA "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.embnet.org/licenses/embnet.html>).
 CC or send an email to license@ebi-ebi.ch.
 CC -----
 DR EMBL; AJ235272; CAA14952.1; -
 DR PIR; F71653;
 DR InterPro; IPR004667; ADP-ATP-car.
 DR Pfam; PF03219; TLC; 1.
 DR TIGRFAMs; TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 512 AA; 58208 MW; D09A265DC38936EB CRC64;

Query Match 33.5%; Score 880.5; DB 1; Length 512;
 Best Local Similarity 36.9%; Pred. No. 1.3e-48;
 Matches 184; Conservative 108; Mismatches 177; Indels 29; Gaps 10;

9 FGKLRSLMFIHTELKVLPMFLMFCITFNVTVDLDTDTLIVGAPGSGAIAIPLK 68
 19 FSKLTVDYIMPIKRIHISKIFLITLMFCILFIQMLIRALKDSIVTWM-IGAEITSPK 76
 69 MLVPCAIIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPTVYPLRDVLH--PTEFA 126
 77 GGVNPSAFILTVITVYKVNRRKANIPYLLISITLTFPALFAYVIFPHEMLHLPVT-V 135
 127 DRLQALIPPGILGLVALIRNTPFAFYVLAELMGSMLSLWNGFANETTKHEAKRYA 186
 136 HNLTRSL-PNLKWFITLLSKMSFSLFYIABLWNVVALLFWQFNNTITVESKPY 194
 187 LFGIGANISILASGRAIWAASKLRASVSEGVDPG---ISRLMLMTIVSGLVMAAS 242
 195 LFGILSGTGYLACHFLENLSNIVYVTKPALOSSFHTLSIQILITVLIGVISKTF 254
 243 YWINKNVLTDP-----RFYNPEEMQKGGKAKPRMNMKOSFLYIDRSPLYLLTLVYAY 297
 255 WLNKRVLDKKNMALLRF-----KTKNSKITAKSFQMLTSRHLRITLILCY 304
 298 GICINLIEVTKWSQKLQYPMNDYSEFMGNSEFNGVSVLIMLFGVGNVIRKFGMLTG 357
 305 GIATNLVEGPMKAAATIKTPTEYAFISYSTGVFTIFVL-LGSNIVRMGFTS 363
 358 ALVTPVWVLLTGLIVFPLVIFRNQASGLVAMF-GTTPMLAVVGAIONILSKTKALF 416
 364 AVITPSIVFTIGLIFFAVNNFEFGAGLIANFILTDPAVAITIGAIQNVLSKSKYTLF 423
 417 DSTEMAYIPLEDOQVKGKAAIDVVAARFGSGGALLIOGLIVIGSIGAMTPYAVIL 476
 424 DSTKEMAYIPLEPIKISGRAADVITGKLGSGSAFLQSLIFILPSASIQSISICLMI 483
 477 LFIIT--AIWLVSAATKLNK 492
 484 IFILTCVTWIMATKLNK 501

RESULT 13
 TLCD_RICPR STANDARD; PRT; 500 AA.
 AC Q05962;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP carrier protein 5 (ADP/ATP translocase 5).
 GN TLCD OR TLCD5 OR RP739.
 OS Rickettsia prowazekii.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.

CC STRAIN=Madrid E; PubMed=9823893;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [2]
 RP SEQUENCE OF 325-500 FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=97419517; PubMed=9274032;
 RA Andersson J.O., Andersson S.G.E.;
 RT "Genomic rearrangements during evolution of the obligate
 RT intracellular parasite Rickettsia prowazekii as inferred from an
 RT analysis of 52015 bp nucleotide sequence.";
 RL Microbiology 143:2783-2795(1997).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AJ235273; CAA15167.1; -
 DR EMBL, Y11778; CAA72457.1; -
 DR PIR, G71633; G71633.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 287 307 POTENTIAL.
 FT TRANSMEM 328 348 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 469 489 POTENTIAL.
 SQ SEQUENCE 500 AA; 57073 MW; FE3DB48D08CF5F72 CRC64;
 Query Match 32.9%; Score 865; DB 1; Length 500;
 Best local similarity 35.7%; Pred. No. 1.2e-47;
 Matches 176; Conservative 109; Mismatches 182; Indels 26; Gaps 10;
 QY 11 KLSFLPPIHTHELKVLPMFLMPCITFNTYVLRDVKDLIYVAPSGAALPFIKFWL 70
 DB 13 KFRALFVPHVNEELGKPIPLSALMFCILFNQNIIRIKDSILISE--ISAEIAGFAKYVC 70
 QY 71 VVPCALIFEMLYATLSNLSKQALFYAVGTFPLFPALFVIVYPLADV--LHTEPRADR 128
 DB 71 VTPVAAAFVILIYAMINHLTFEKFYYSAFISCFILFAFVYIPNHIHFVHDITLSDW 130
 QY 129 LQALIPGELGLVAILRNWTFPAFYVLAELGWSVLSLMPFGANETIKIHEAKRYALF 188
 DB 131 MNKY--EHEKRYILVNGWGIYVYSLAELMPNIFVULFQFNEITTTBEAKRFTILP 188
 QY 189 GIGANISLSAGRIWASKIRASVSECDPWGISLRLMAMTVSGLV-----LMSYV 243
 DB 189 SLFNSSLIIVGFLMMNISSEDTIIRKPISSDSKITLVQSTIIAIVAILCCILVRF- 247

QY 244 WINKVLTDRPFPYNEEMOKKGAEPKMMKDSFLYDSSPYLTLTLVAVGICINL 303
 DB 248 -ISKYITNLPFH--KTSSTSTAPRMGLISFKYIVASKYIMLLIGSAFPAINL 303
 QY 304 IEVTKSQQLQYPPNNNDYSEFMGNFSGFTGVSVLIMLFGVGNVIRKFGMLTALVTPV 363
 DB 304 VEAVWKAKIKELVPTVTVAFNLSVILMTG-VAIIMTIIIGNVMMHMFVAIVISPV 362
 QY 364 MVLITGVFPALVIFRNOA---SGLVAMFGTTPMLAVVGAIONLSKTKALDST 419
 DB 363 IIMVGVLPFGILVFPDQIISLFDGALIM--SPLAAVSIIGIIONLAKGTYSIDTS 419
 QY 420 KEMAYIPLDOEKYKGAADIVVAARFKSGGALIOGGL--LVYCGSIGAMTPYLAITL 477
 DB 420 REMLYIPLDELTKKGAAVDIVSAKVGKSSGLVQSIIFTLVNNATPSTISPLMVVFT 479
 QY 478 FIAIMLVSAATKL 490
 DB 480 FVCGAMIVAVRKI 492
 RESULT 14
 ID TLCB_RICPR STANDARD; PRT; 507 AA.
 AC 092DF2;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP/ATP carrier protein 2 (ADP/ATP translocase 2).
 GN TLCB OR TLC2 OR Rp377.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxId=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN
 CC EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.
 CC THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF
 CC RICKETTSIAL PARASITISM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ADP/ATP TRANSLOCASE TLC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AJ235271; CAA14836.1; -
 DR PIR, B71695; B71695.
 DR InterPro: IPR004667; ADP_ATP_car.
 DR Pfam: PF03219; TLC; 1.
 DR TIGRFAMs: TIGR00769; AAA; 1.
 KW Transmembrane; Transport; ATP-binding; Multigene family;
 KW Complete proteome.
 SQ SEQUENCE 507 AA; 58457 MW; 00DB2DFCC51D9D59 CRC64;
 Query Match 31.3%; Score 824; DB 1; Length 507;
 Best local similarity 34.2%; Pred. No. 4.5e-45;
 Matches 175; Conservative 114; Mismatches 170; Indels 52; Gaps 10;
 QY 11 KLSFLPPIHTHELKVLPMFLMPCITFNTYVLRDVKDLIYVAPSGAALPFIKFWL 70

```

Db 18 KFRIVWPVRSYELTKFIPMLMFILLNINVRISDSVVLISS--EVLSPILKMG 75
Qy 71 VVPCALIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPVIVYPLRDVLPTEPADRLQ 130
Db 76 EMPGVFVLIYSKLCNMTTEQVFRITSTFFLFFALFGLPPYKEFFHNPBELLNOY 135
Qy 131 AILPGLGLVAILRNMTFPAFVYLAELMGSMVLSIMFGANETTKYHEKRFYALFGI 190
Db 136 IIVPLHKLWFLIWMQNSLVLYTIMGELMPVIVFLWQJLANKITVKEBAPRYFSFTL 195
Qy 191 GANISLASGRAVWASK-----LRASVSGVDPMGDISRLMAMTVISGLVLMASVW 243
Db 196 FQGNILFSGVLIYIPAKSEHFLPLFAHLN---DTNELLSKFTVILLISGLCLMLHK 252
Qy 244 WINNVLTDPFPYVPEMOKGKGAKP-----KANKDSFLYDRSPYILLTLVAYG 298
Db 253 LIDKSVV-----EADKNIKFKRNORTDILKLSAKIILTSRGLGICLLVMSYS 303
Qy 299 ICILILEVTWKSOLKQVPMNDYSEPMGNFSPWTVGVSVLMLFVGGNVIRKFGMLTGA 358
Db 304 MSNVLISGLMWSKVKOLYPATKDFISIHGEVLFWTGVLT-LVSAFLSGSLIRIYGMWGA 362
Qy 359 LVPVWVLLTGVFPAVIFPNQASGLVAMFG-TTPLMLAVVGAIONILSKSTKVALFD 417
Db 363 IITPIMFVAVGMFFSPFIHQHGNIVNTLGSSPLVIVIFIGMLVAFKASGLSFD 422
Qy 418 STKEMAYIPLDOEKVKGKAAIDVVAARFGSGGALIO-----OGLVYC 462
Db 423 ATKEMVYIPLDNEIKTKGKAADVWGAKIGKSGIILIOFISFIFPNVHNDIAGL--- 479
Qy 463 GSIGAMTPYLAIVLLEIATVLSATKLNKL 493
Db 480 -----WTFPIVCIIMLYGVKVL-SQYTKM 504

```

```

RESULT 15
LACY LEULA
ID LACY LEULA STANDARD; PRT; 639 AA.
AC Q48624;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactose permease (Lactose-proton symport) (Lactose transport protein).
GN LACS.
OS Leuconostoc lactis.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_Taxid=1246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZ6009;
RX MEDLINE=96209221; PubMed=8633855;
RA Vaughan E.E., David S., de Vos W.M.;
RT "The lactose transporter in Leuconostoc lactis is a new member of the
RT Lacs subfamily of galactoside-pentose-hexuronide translocators.";
RL Appl. Environ. Microbiol. 62:1574-1582(1996).
CC -1- FUNCTION: RESPONSIBLE FOR TRANSPORT OF BETA-GALACTOSIDES INTO THE
CC CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM),
CC AND ALSO FOR TRANSPORT OF HOMOLOGOUS AND HETEROLOGOUS EXCHANGE OF
CC BETA-GALACTOSIDES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DOMAIN: THE ENZYME IIA-LIKE REGION MAY SERVE A REGULATORY
CC FUNCTION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC SODIUM-GALACTOSIDE SYMPORTER FAMILY (SGP).
CC -1- SIMILARITY: Contains 1 Pts EIIA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/> or send an email to license@ebi-sib.ch).

```

CC -----
DR EMBL; U47655; AAC44113.1; -.
DR HSSP; P45618; ZGPR.
DR InterPro; IPR001927; Na/Gal symport.
DR InterPro; IPR001127; PTS EIIA.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR TIGRFAMs; TIGR00792; gph; 1.
DR TIGRFAMs; TIGR00830; PTBA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR PROSITE; PS00872; NA_GALACTOSIDE_SYM; 1.
KW Transmembrane; Sugar transport; Transport; Symport; Phosphorylation;
KW plasmid.
FT DOMAIN 1 473 SGP DOMAIN.
FT DOMAIN 474 639 EIIA DOMAIN.
FT MOD_RES 557 557 PHOSPHORYLATION (BY HPR) (BY SIMILARITY).
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
SQ SEQUENCE 639 AA; 70153 MW; 25DF2819761B415 CRC64;

```

Query Match 5.2%; Score 135.5; DB 1; Length 639;

Best Local Similarity 20.0%; Pred. No. 0.13; 209; Indels 129; Gaps 22;

Matches 104; Conservative 78; Mismatches 209; Indels 129; Gaps 22;

```

Qy 54 GAPSGAEAPVIFKFMVLPVCAIIFMLIYAKLSNLSKQALFYAVGTPPLIFPALFPPTVI 113
Db 49 GLPQSVANKILGLITLVAIIRLAVIDILGNIVNTNRNGKPFQWVIGAVSSV- 107
Qy 114 YPLRDVLPTEPADRLQALIPPLGLGLVAILRNW-----TPAAFYLAELMGSMVLSMF 168
Db 108 -----LLVVIPTGIFGLAHI--WIAFAIVFVLFIILDFYS-FADVAY 149
Qy 169 WGFANEITKIHAKRFALFG-----IGAN-----ISLASGRAVWASLTRA 211
Db 150 WGVVPALISDSKSGKGIPTSGSFTGSGIMNGLTMIIVPVTYTFYIAFGK- 200
Qy 212 SVSEGVDPW-GISRLMAMTVISGLVLMASVWVINKNVLTDPFPYVPEMOKGKGAKP 270
Db 201 --EGGPGWGFGL-VVSIYAVSALA-----AFGRKEDNIRNATK 242
Qy 271 KANKDSFLYDRSPYILLTLVIAVGINLIE--VTWKSQKQVPMNDYSEPMGN 328
Db 243 KTSIKDVFSGIINHDIIMSLAVMYSLAYVVTNGVLPFFKFKVLGKPN----- 292
Qy 329 FSPFTGVSVLMLFVGVGNVIRKPGMLTGALVTPVWMLLGIYFPAVIFRNQASGLVAM 388
Db 293 -EFV-----IAGALVYIGFSTAPLVPLNFKTRKQVLSGQAMILSYLFTI 340
Qy 389 FGTPMLAVVGAIONILSKSTKVALFDSTKEMAYIPLDOEKVKG-----KAIDVVA 443
Db 341 FGTNMM--VTIGILNFTFAQLVVLSTLDSEYQGLKNGRNEAVVLAVRPMIDKIT 399
Qy 444 ARF--GKSGALIQOGLL--VICSGIGA-----MTPYLAIVLFIATVLSAT 488
Db 400 GAFNGVAVGAITAGMTGATAGDISASKINTFEIYAFYPLFLSILALVILMKVKIT 459
Qy 489 -----KLNKFLAQSALK--EGEV---AQEDSAPAS 514
Db 460 EKKAELVIELKT-LSSGAKKANTSEVNVLELEIFAPAS 498

```

```

RESULT 16
YH06_HAEIN STANDARD; PRT; 669 AA.
ID YH06_HAEIN
AC P45335;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable transport protein H11706.
GN H11706.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BCCT (TC 2.A.15) FAMILY OF
CC TRANSPORTERS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U3843; AAC23352.1; -.
DR PIR; D64137; D64137.
DR TIGR; H11706; -.
DR InterPro; IPR000060; BCCT transporter.
DR Pfam; PF02028; BCCT, 1.
DR ProDom; PD010111; BCCT transporter; 1.
DR TIGRFAMs; TIGR00842; bcct; 1.
DR PROSITE; PS01303; BCCT, 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 259 279 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 470 490 POTENTIAL.
SQ SEQUENCE 669 AA; 75020 MW; 0AEE803BB5C733F CRC64;
Query March 5.1%; Score 133; DB 1; Length 669;
Best Local Similarity 19.8%; Pred. NO. 0.19; Indels 216; Gaps 30;
Matches 116; Conservative 87; Mismatches 167;
15 FLNPIHTEH-----KVLPMFLMFCITFNTVLRDVKTLIAGAGSGAALPFIKFW 69
27 FIAEGQALINQAKSGIFANFSFWYVITFSV-----FAGFL 63

```

```

QY 70 LVVPCALIFMLIYAKLSNI-----LSKALFYAVGTPF-LIFPALPFTVTP 115
DB 64 LI-----LSVSSIGNIKLGDDEEPEPFSLSMLMLPAAGMGVGFVAEPITTH 115
QY 116 LRDVLPTEPADLQAILPPGLIGLVAIIRMTFAFYVLAELGSMVSLMFWGANET 175
DB 116 LSDITGASAHKQOQELL-----HTLFPHGIIHMAV--YGTALALAVGFPRYL 163
QY 176 T-----KHEAKRYA---LFGI-----GANISLASG-RAIWAASKL 209
DB 164 PLALRSCFYPILKDRINGKIGDAIDWALATLFGIITTLGFOSSQIGALDQIGWLSQW 223
QY 210 RASVSEGVDPWGISLRLMAMTIVSGVLMASYVMWINKVLTDRFNPBEMOKGKGAK 269
DB 224 SPALQGVLT--VVMCLAVSAISGV-----GKGVK 252
QY 270 --PRMMKDSF---LYLDRSPYILLTLTVIAYG-ICIMLEIYVWKSQKLQYPMNDY 322
DB 253 ILSEINLTAFCLLPVLSGPTLYLSAIFSDNIGVFSVLVQLSFT-----YAYQEH 307
QY 323 SEFMGNFSF-----WGVSVLIMLVGVNIRKFGMLTGALVTPVNVLLTGIVFPA 374
DB 308 TSWFSGWTVMYAMWCSWAFVGLFIARISKGRITREF--IFGLVIP---SLFGILMF- 361
QY 375 LVIFRQASGLVAMFGTTPML--AVVGAIQWILSKSTKALPDSTKEMAYIPL----- 427
DB 362 -----TVFGNTAVLNDGIAAGLGEPIS-SPILLF--KPLNYLPLPTTG 405
QY 428 -----DOE-----QKVGKRAIDVVAAREFGSGG-A 452
DB 406 FVSLVILVLFITSADSGIYVANNIASRDKSLASPAKQALIMKGTMSVVAIVLMQSGLA 465
QY 453 LIQGGILVIGSGIGAMTPYLAIVLFIATW--LVS-----ATKLN 491
DB 466 NLQTMTLIVA-----LFPALLMLVMCFSLMKGLADKKYFSTKVN 505

```

```

RESULT 17
NARK_ECOLI STANDARD; PRT; 463 AA.
ID NARK_ECOLI
AC P10903;
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrite extrusion protein 1 (Nitrite facilitator 1).
GN NARK OR B1223.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89338707; Pubmed=2668029;
RA Naji S., Nohno T., Saito T., Taniguchi S.;
RT "The nark gene product participates in nitrate transport induced in
RL Escherichia coli nitrate-respiring cells.";
RN FEBS Lett. 252:139-143(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1232-1244(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; Pubmed=8905232;

```

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA itemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.,
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=9251357; PubMed=9097039;
 RA Nobuo T., Noji S., Taniguchi S., Saito T.,
 RT "The *narX* and *narJ* genes encoding the nitrate-sensing regulators of
 RT *Escherichia coli* are homologous to a family of prokaryotic two-
 RT component regulatory genes.";
 RL Nucleic Acids Res. 17:2947-2957(1989).
 CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRITE PRODUCED BY THE
 CC DISIMILATORY REDUCTION OF NITRATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (probable).
 CC -1- INDUCTION: BY NITRATE UNDER ANAEROBIC CONDITIONS.
 CC -1- MISCELLANEOUS: SINCE THE TRANSPORT IS CONTROLLED BY OXYGEN, FIVE
 CC CYS RESIDUES IN *NARX* MAY PLAY AN IMPORTANT ROLE IN SENSING REDOX
 CC POTENTIAL TO REGULATE THE FACILITATOR.
 CC -1- SIMILARITY: BELONGS TO THE *NARX*/NARX FAMILY OF TRANSPORTERS.
 CC -1- SIMILARITY: STRONG, TO E. COLI *NARU*.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X15996; CAA34126.1; -
 DR EMBL: X69189; CAA48933.1; -
 DR EMBL: AB000220; AAC74307.1; -
 DR EMBL: D90757; BAA36091.1; -
 DR EMBL: D90758; BAA36093.1; -
 DR EMBL: X13360; CAA31740.1; -
 DR PIR: S05239; GRECON.
 DR Ecocore, EG10642; *narX*.
 DR InterPro: IPR004737; NO3 transporter.
 DR TrEMBL: TIGR00886; 2A0108; 1.
 KW Nitrate assimilation; Transport; Transmembrane; Inner membrane;
 KM Multigene family; Complete proteome.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 79 99 POTENTIAL.
 FT TRANSMEM 104 124 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 178 198 POTENTIAL.
 FT TRANSMEM 212 232 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 320 340 POTENTIAL.
 FT TRANSMEM 345 365 POTENTIAL.
 FT TRANSMEM 405 425 POTENTIAL.
 FT TRANSMEM 436 456 POTENTIAL.
 SQ SEQUENCE 463 AA; 49693 MM; AF1D67CAD40FE8A CRC64;
 Query Match 5.0%; Score 132.5; DB 1; Length 463;
 Best Local Similarity 22.3%; Pred. No. 0.14;
 Matches 108; Conservative 69; Mismatches 174; Indels 133; Gaps 28;
 Oy WLVPCALIMLYAKLSNI-LSKQALFYANGTFFLLFPFVIVPLDVLHPTFEAD 127
 Db WISPCCLLAFPCWMLFSAAVAVNPKVGFNTTQQLFNLTRALPVSQALRV--PYSF-- 94
 Oy 128 RLQALIPGLGLGVALI--RNWTFAPAFVLAELMGVYMLSMFPGFANEITKIHAEAFY 185

Db 95 -----MWPIFGRRWT--AFST-----GILLIPCVWLGF-----VQDISTPY 130
 Oy 186 ALGIGANISLLASGAIWASKLRASVS-----EGVDPWGISRLMM 230
 Db 131 SVFII--ISLLGFGAGANASSM-ANISFFPKQOGALGLNGGNGSVMOVLAP 186
 Oy 231 TIVSGVLVMSYMWINKVLTDRFYNPEMQKKKAKPKMMKDSFLYLDSPYLL- 289
 Db 187 LVYS-LSIFAVF-----GSGVQKPDGTE--LVLANSWITWP 221
 Oy 290 -LTLVLVAVGICINLIEVTKSQLQYPMND-----YSEKGNF-SFWTV-- 335
 Db 222 FLAIFTLAAMFGNMDL-ATSKASIKEDPLVKRGLHIMSLVLTATPGSIFGSAGPML 280
 Oy 336 -----VSLVILMPFG--GNVIRKFGW-----LGLALVTPV-----MVLTVGFVALV 376
 Db 281 SKTQFPDVQILOYAFQFPFGLGALARSAGALSRLGRTVLVFLIMAFSGULFTLP 340
 Oy 377 IFRNQSGLVAMFGTTPMLAVV-----GAIONILSKSTKXALPSTKEMAVIPDOE 430
 Db 341 T-DQGGSPMAFPA--VFALFLTAGLSGSGTFQMISVFRKLTMDRVAAEG--GSD 392
 Oy 431 QYKGAADIVAA-----RFGSGGALLQOGL--LVIGGS-IGAMTPYLAIVLLFIIA 481
 Db 393 ERAMREKATDTAALGFIASIGAFGFPFKAFGSSIALTGSVGMKVLIFVIACVI 452
 Oy 482 TWLV 485
 Db 453 TWAV 456
 RESULT 18
 ID NARU_ECOLI STANDARD; PRT; 462 AA.
 AC P37758; P77696;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nitrite extrusion protein 2 (Nitrite facilitator 2).
 GN NARU OR B1469.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Bonney V., Ratouchniak J., Biasco F., Chippaux M.,
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blactner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitagawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
 RA Tagami H., Takeda U., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horuchi T.,
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).

FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 FT TRANSMEM 464 484 POTENTIAL.
 SQ SEQUENCE 512 AA; 56122 MW; FID8B1EEB30F1FF CRC64;

Query Match 4.8%; Score 126.5; DB 1; Length 512;
 Best Local Similarity 19.5%; Pred. No. 0.38;
 Matches 109; Conservative 82; Mismatches 160; Indels 201; Gaps 27;

QY 68 FVLV-----PCAIIMFLIYA-KLSNI-LSKQAL-----FYAVGNPFLIFALPFTVYYP 115
 DB 9 FVVIATTAALVAVGVSPLDQVNSQAQAFIDBSFQVILVLSFVGCLF-LIFSP 67
 QY 116 LRDVL-----HPTPADRLQAILPGLIGLVAIIRMTFAFYVLAELMSVMSLMFWG 170
 DB 68 IGKIKLGPDKPEPF-----GLLSWFAML-----GMSGIGLVFYG 104
 QY 171 FANBI-----TKIHEAKRYALFGIGANI-----S 195
 DB 105 AAEPIISHVAISSPGSETTPOAFRDALRYTFEHHGLHMAIYAIVALCIAVFOFKGAPG 164
 QY 196 LLSAGRAIWMASKLRASVSEGV-----PMGISLRLLMAMTIVSG----- 235
 DB 165 LISSTLSPILDKXNGPIGKALDCLAVPAIVGVSTSLGATQINGANLFGIPNAFI 224
 QY 236 -----LVMAIYMWINKNVLTDFRYNPEEMQKKGAKPKNMKDSFLYLDSP 285
 DB 225 VQVLIIIVVLFILISAM-----SGLGKGIK-----YLSNTN 256
 QY 286 YI-----LLTLTIVANGICI-----NLIEVYMSQKLQYPMNDYSEMGN 328
 DB 257 MVLAGLMLLELVGPVLLINNSTDSIGQYIQNIIVQMSFPLT-----PNDPEKREWINS 311
 QY 329 FSF-----MTGVSVLIMLFVGNVIRKFGMLTGALVPPVWLLTGIYFPALVIFRN 380
 DB 312 WTIFFYMAWMTSMSPFVGIFIAVSRGRITREF-LIGVLVTPCLITELMPSIFGVASMDL 369
 QY 381 QASGL-----VAMFGTT--PL-----MLAVVGAIONILISKTKALPDSTKEMA 423
 DB 370 QOKGAFVNAKLSTETMFGTLIDHYPLIMVTSILAILIHAVFISADS--ATFVLGMQTS 427
 QY 424 YIPLDQOKYKKGKAI--DVVAARFGSGGALLIOGGLVIGSGANGTPYLAIVLLITIA 481
 DB 428 YGSLNPANSVYLSNGIIOQAMAAVLLVSGGLAALONTAIL-----AALPFSIVILLMIAS 482
 QY 482 IW-----LVSATKLNK 492
 DB 483 LYOSLSKEREREIKAKELDK 502

RESULT 21
 ID COX1 CYACA STANDARD: PRT: 526 AA.
 AC P48867;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1 OR COX1.
 OS Cyandidium caldarium.
 OC Mitochondrion.
 OC Bakaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Cyandidium.
 OX NCBI_Taxid=2771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RK-1;
 RA Viehmann S.;
 RL Theiss (1995), Justus Liebig University / Frankfurt, Germany.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY

CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----
 CC EMBL, Z48930; CA88773.1; -;
 CC DR PIR, S62763; S62763.
 CC DR HSSP, P98002; IARL.
 CC DR InterPro, IPR000883; COX1.
 CC DR Pfam, PF00115; COX1; 1.
 CC DR PRINTS, PR01165; CYCOXIDASE1.
 CC DR PROSITE, PS00077; COX1; 1.
 CC KX Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 CC KM Respiratory chain; Inner membrane.
 CC FT METAL 65 65 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 CC FT METAL 244 244 COPPER B (PROBABLE).
 CC FT METAL 248 248 COPPER B (PROBABLE).
 CC FT METAL 293 293 COPPER B (PROBABLE).
 CC FT METAL 294 294 COPPER B (PROBABLE).
 CC FT METAL 379 379 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 CC FT METAL 381 381 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 CC FT CROSSLINK 244 248 1'-histidyl-3'-tyrosine (By similarity).
 CC SQ SEQUENCE 526 AA; 56354 MW; 5A6CA16E336901B6 CRC64;

Query Match 4.8%; Score 126.5; DB 1; Length 526;
 Best Local Similarity 18.4%; Pred. No. 0.39;
 Matches 89; Conservative 70; Mismatches 160; Indels 165; Gaps 21;

QY 19 IHTHELKV-----PMLFFCIT-----FNYVLRDTKTLIYGAPSGAEALPIFK 67
 DB 52 IGHQLVNVLVTAHGLILFFVVIPTLMGGFGNMFV-----PLIIGAPMAAPRLNNIS 105
 QY 68 FVLVPCALIFM-----LIYAKLSNLSKQALFYAVGTFELFPLFPTVY 114
 DB 106 FVIMPPSLIILLSAFVETGAGTWLTPPLSSV--QA--HSGGAVDLAIFSLH--IS 157
 QY 115 PLRDVLHPTFEADRLQAILPPG-----LLGLVAIL--R 145
 DB 158 GISSILGASNFATITNIRNPQGNLVRIPLFVSVLVTAIILITLPVLAGATITLLTDR 217
 QY 146 NMTFAFYVLAELMSVMSLMFW-----GFANEITKHEAKRYALFG- 189
 DB 218 NFN-TSFFDSGGADPVLFGHLFWFGHPEVYILVPAFGIISQVSTFSRKKVFGVGI 276
 QY 190 IGNISILASGRAIV-----ASTLRASVSEGV-----DPWGISL 224
 DB 277 IYALISIRILG-SVVAHMHMTTGMDVTRAYTAAASLIAVPTGIVKSVWATMWGSI 335
 QY 225 RLIMANTIVSGLVMAIYMWINKNVLTDFRYNPEEMQKKGAKPKNMKDSFLYLDPS 284
 DB 336 SLKTPMLFAIGFIILFTVGLTGLVAVANSGL-----DISLHDYVVAHF 380
 QY 285 PYILLTLVIANVCITNILEVTWKSQKLQYPMNDYSEMGNSFQVGVSVLIMLF- 343
 DB 381 HYVLISGAL--GFIFAGF--YYWIKI-----CGKQYSETLIQIHFMTIFIGVNLTFPP 430
 QY 344 -----VGNVIRKFGWMLTGALVTPVWVLLTGIYFPALVIFPNQAS 383

Db 431 MHFGLAGIPRIIPDPAYEEMNIVSTRG-----AKSILGTLFFVYVLAFTN 481
 QY 384 GLVA 387
 Db 482 GLIS 485

RESULT 22
 ID PTSA STRMU STANDARD; PRT; 664 AA.
 AC P12655;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PTS system, sucrose-specific IIAB component (IIAB-C-SCR) (Sucrose-
 permease IIAB component) (Phosphotransferase enzyme II, ABC
 component) (EC 2.7.1.69) (BII-SCR).
 GN SCRA OR SMU.1841.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus;
 OK NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS-5;
 RX MEDLINE=89123027; PubMed=2536656;
 RA Sato Y., Foy P., Jacobson G.R., Kuramitsu H.K.;
 RT "Characterization and sequence analysis of the scra gene encoding
 RT phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
 RL J. Bacteriol. 171:263-271(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186;
 RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 RT pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 RN [3]
 RP SEQUENCE OF 639-664 FROM N.A.
 RC STRAIN=GS-5;
 RX MEDLINE=93329360; PubMed=8336109;
 RA Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
 RT "Isolation, characterization and sequence analysis of the scrc gene
 RT encoding fructokinase of Streptococcus mutans.";
 RL J. Gen. Microbiol. 139:921-927(1993).
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.
 CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
 CC histidine + sugar phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Contains 1 PTS EIIA domain.
 CC -1- SIMILARITY: Contains 1 PTS EIIIB domain.
 CC -1- SIMILARITY: Contains 1 PTS EIIC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M22711; AAA26971.1; --

DR EMBL; AE015011; AAN59464.1; --
 DR EMBL; D13175; BAA02466.1; --
 DR PIR; B32243; B32243.
 DR HSSP; P08637; IGIC.
 DR InterPro; IPR001127; PTS_EIIA.
 DR InterPro; IPR001966; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIA_1; 1.
 DR Pfam; PF00367; PTS_EIIB_1.
 DR Pfam; PF02378; PTS_EIIC_1.
 DR Pfam; PF002243; PTS_EIIA_1.
 DR Pfam; PF001476; PTS_EIIB_1.
 DR TIGRFAMs; TIGR00826; EIIB_GLIC_1.
 DR TIGRFAMs; TIGR00830; PTBA_1.
 DR PROSITE; PS00371; PTS_EIIA_1; 1.
 DR PROSITE; PS01035; PTS_EIIB_CYS_1.
 KW Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Transmembrane; Complete proteome.
 FT DOMAIN 1 40 EIIB DOMAIN.
 FT DOMAIN ? ? EIIA DOMAIN.
 FT DOMAIN 533 664 EIIA DOMAIN.
 FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 664 AA; 69988 MW; 809E63E32281A9A1 CRC64;

Query Match 4.8%; Score 126; DB 1; Length 664;
 Best Local Similarity 19.9%; Pred. No. 0.53; Indels 200; Gaps 26;
 Matches 112; Conservative 83; Mismatches 168;

QY 16 LMPHITHELVKLVLMFLMFCITFNYTVLRDTRKDLIVGARGSGAEAFPIKFWLVPCA 75
 Db 82 LTVSTDDDKKIASGKGFPMALITLSDI-----FPIIPAL 121

QY 76 IIFMLYAKLSNLSKQALFYAVGTPFLIFPAEFTVIYPLRDVLPHTPADRIQALIP 135
 Db 122 VAGGLVMA-LNNPLTSEGLF---GTSILV--QQPPII-----KGSSDMIQ----- 160

QY 136 GLGLVAILNNWTAATVYVLAELMGVMSLMFGEFANEITKIHAEKFPALPFGIGANTIS 195
 Db 161 -----LMSAPFW-FLPITVGSAAKFGANQFLGASIG 193

QY 196 LL--ASGRAIWMASKIRASVSEGVDP-----WGI-SLRLLMNTIVSGVLMSYWMIN 246
 Db 194 MIMVAPGANINIGLAANAIPISKATIGATYGFMNIRBELHTQASVYTVQVPLVAVVLL- 252

QY 247 KNVLTDPRFYNPBEEMQKKGAKPKMMKDSFLYLDKSPYL--LTLTVIANGICINLI 304
 Db 253 -SIL--EKFFH-----KRLPSAVDFEFTPLSLVITNGFLTFIVIG----- 289

QY 305 EVTMKSQKIQPYMNDYSEFMGNFSW----TGVSVLMLPFGNGVIRKFGMLTGALV 360
 Db 290 -----PVMKVSQWLTNGIWLVDYDTG-----FLGKGVF-----GALY 322

QY 361 TPVWVLLTGI--VEFALVIFRNQASGLVAMF--GTPPLMLAVVGAIONLSKTKYALF 416
 Db 323 SP--VWVTGLHSGSPAL-----ETQLISAFOQNTGSGDIFPYASMANVAAQGAATPAIV 374

QY 417 DSTKEMAYIPLDQKQYKGAIDVVAARFGSGGALIQGGL---VIGSIGAMTPYL 472
 Db 375 FLTK-----DKMKKGSSSGVAGALIGTEPALFGVNLKXRPFFCALIGSASAAA 425

QY 473 AVILLFIATWVSA-----TKNKFLA 496
 Db 426 IAGLLQVAVASLSAGLGLSLKASSIPYVVCCELISPAIAVAVYGYGKTYAVDVFAA 485

QY 497 QSALKE-----QEVQEDSAPAS 514
 Db 486 EAAVEAIEVQETPEBAASAN 508

RESULT 23
 RAEP_PEDPE


```

ID  RAEP PEDPE  STANDARD;  PRT;  641 AA.
AC  P43466;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Raffinose carrier protein (Raffinose permease).
GN  RAEP.
OS  Pedicoccus pentosaceus.
OC  Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pedicoccus.
OX  NCBI_TaxID=1255;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=PEP1.0;
RA  Ienhouits K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL  Submitted (xxx-1994) to the EMBL/Genbank/DBJ databases.
CC  -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC  -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC  SODIUM:GALACTOSIDE SYMPORTER FAMILY (SGP).
CC  -1- SIMILARITY: Contains 1 Pts E11A domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z32771; CAAB3664.1; -
DR  EMBL; L32093; AAA25663.1; -
DR  PIR; S44253; S44253.
DR  HSSP; P20166; 1GPR.
DR  InterPro; IPR001927; Na/Gal symport.
DR  InterPro; IPR001127; Pts_E11A.
DR  Pfam; PF00358; Pts_E11A_1; 1.
DR  ProDom; PD002243; Pts_E11A; 1.
DR  TIGRFAMs; TIGR00792; Gph; 1.
DR  TIGRFAMs; TIGR00830; PTBA; 1.
DR  PROSITE; PS00371; Pts_E11A_1; 1.
DR  PROSITE; PS00872; NA_GALACTOSIDE_SYMPT; 1.
KW  Transmembrane; Sugar transport; Transport; Symport; Phosphorylation.
FT  DOMAIN 1 641  E11A DOMAIN.
FT  MOD_RES 559 641  PHOSPHORYLATION (BY HPR) (BY SIMILARITY).
FT  TRANSMEM 25 45  POTENTIAL.
FT  TRANSMEM 57 77  POTENTIAL.
FT  TRANSMEM 93 113  POTENTIAL.
FT  TRANSMEM 120 140  POTENTIAL.
FT  TRANSMEM 168 188  POTENTIAL.
FT  TRANSMEM 201 221  POTENTIAL.
FT  TRANSMEM 253 273  POTENTIAL.
FT  TRANSMEM 288 308  POTENTIAL.
FT  TRANSMEM 317 337  POTENTIAL.
FT  TRANSMEM 342 362  POTENTIAL.
FT  TRANSMEM 394 414  POTENTIAL.
FT  TRANSMEM 429 449  POTENTIAL.
FT  TRANSMEM 485 505  POTENTIAL.
SQ  SEQUENCE 641 AA; 69913 MW; 15BEC6F9FC0F61 CRC64;

Query Match 4.8%; Score 125.5; DB 1; Length 641;
Best Local Similarity 19.1%; Pred. No. 0.55;
Matches 92; Conservative 83; Mismatches 156; Indels 151; Gaps 24;

```

```

DB  188  --FSAKNTSGDNGWIFALICILALISA-----NGVGLG--- 222
QY  229  AMTVSGLVLAWSWINKNTLTPRFNPEBEMQKKGAKPKANMDSPYLDRSPYL 208
DB  223  -----TREYDSDIR-----KNKQDTGVMEIFALANDDL 254
QY  289  LTLTLVAVGICINLIEVTWKSOLKLOYPNNDYSEPMGN-----PSFMGVSVL 339
DB  255  WAAALVIFYGVGINL-----GSLFVY-----FTYIMGRKKSILSTINIFGLIRTS 304
QY  340  IMLFVGVNIRKFGMLTGAALVTPVWVLTGTIVFPAALVIFRNOASGL-----VAMFGTTP 393
DB  305  LFPVLSSKFSRK-GVPAQCIV-----FMLGIAIFTIA-----GSNLMVLAAATWFGFPQ 354
QY  394  LMLAVVGAIGNILSKTKYLF-----DSTKEAVIPL-DQEQVKGAALDVAAAFG 447
DB  355  QMFLVAV---LMVITDSVEYQQLKGRHDESLASVPLIDFGAISNGVAGQIAIISG 411
QY  448  KSGGALIQGGLLVICSGIG-AMTPYLAIVLIFIAIMVLSATKINKFLAQSALKEQVVA 506
DB  412  MTTGATLASS--ITTAQQUHFKLTWAPPALMLTALIGIPS---KQIFLTEE--KHARIV 463
QY  507  QE 508
DB  464  AE 465

RESULT 24
YWF_BACSU
ID  YWF_BACSU  STANDARD;  PRT;  394 AA.
AC  P39642;
DT  01-FEB-1995 (Rel. 31, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Hypothetical protein ywf.
GN  YWF OR IPA-84D.
OS  Bacillus subtilis.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=1423;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE=95020537; PubMed=7934828;
RA  Glaser P., Kunz F., Arnaud M., Coudart M.P., Gonzales W.,
RA  Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA  Pressecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA  Rapoport G., Danchin A.;
RT  "Bacillus subtilis genome project: cloning and sequencing of the 97
RL  Mol. Microbiol. 10:371-384 (1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE=98044033; PubMed=9384377;
RA  Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA  Azevedo V., Berteiro M.G., Besieres P., Botochin A., Borchert S.,
RA  Borris R., Boursier L., Brans A., Braun M., Briquet S.C., Bron S.,
RA  Briouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA  Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA  Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA  Eutican K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA  Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galletton N.,
RA  Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA  Giuseppe G., Guy B.J., Haga K., Hache J., Harwood C.R., Henaut A.,
RA  Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA  Joris B., Karamata D., Kaasahara Y., Kleier-Blanchard M., Klein C.,
RA  Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA  Kurita K., Lapius A., Lardinois S., Leuber J., Lazarevic V.,
RA  Lee S.W., Levine A., Liu H., Maeda S., Mauel C., Medigue C.,
RA  Medina N., Meliando R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,
RA  Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea B., Park S.H.,
RA  Paro V., Pohl T.M., Portetelle D., Portolick S., Prescott A.M.,
RA  Pressecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

```

RA Rieger M, Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Setiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sotokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuti R., Wedler E., Wedler H., Wellenzeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Zumslein E.,
RA "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT subtilis".
RL Nature 390:249-256(1997).
CC - SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcement/>
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL, X73124; CAB51640.1; -
DR EMBL, Z99123; CAB51797.1; -
DR PIR, S39739; S39739.
DR Subtilist; B010630; ywF.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 POTENTIAL.
FT TRANSMEM 31 POTENTIAL.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 142 162 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 353 373 POTENTIAL.
SQ SEQUENCE 394 AA; 43416 MW; CFE8B079F0A56A98 CRC64;

Query Match 4.7%; Score 123.5; DB 1; Length 394;
Best Local Similarity 20.4%; Pred. No. 0.45;
Matches 98; Conservative 77; Mismatches 166; Indels 139; Gaps 22;

QY 61 EAIPPIKFWLVPICAIIMLIYAKSNILSKQALFYAVGTPPLIFPALFPYIYPLRDYL 120
DB 15 QALSFMGQCYVLPAALLISTYHD-----YMW-TSGVIVRSIIPWFGPFLGVL 62
QY 121 HPTFADPLQAI-----LPPG--LGLVALIRKWTAAFPVLAEL 158
DB 63 -----VRLDRIKIMLWTDIIRGIIELGUTLPKGEYPLIFALLFTYGGVFENPARL 117
QY 159 WGSVWLSLMEFGFANEITKIHAKRFVALFGIANGISLASHGALVWASKLRSVSEGV 218
DB 118 --AVMSLE-----SDISINTLFAKATITISIVGAAA-----GGLF 152
QY 219 PWGISLRLMAMTIVSGVLMAVYWKINKVLTDERFNPBEMOKGKGAAPKMMKDSF 278
DB 153 LLGGSVELAVAF--NGVTIVLSAFIRIKLQ--FVPIQSEIRKFAFQ--SFEKGL 202
QY 279 LYDRSPYIL--LTLVLVIA--YGCICMLIEVTKWSQKLQYPPNNNDSEFMGNSFWTG 334
DB 203 KEITNSNVLNAMFMTIMTALMLGVAVYFPV-----SRPLGDEIGNF 247
QY 335 VSVVLVIML--FVGGVIRKFGMLTGCALTVPVWVLLTGIVFPALVFRNQASGLVAMFGTY 392
DB 248 ILTFCIGGGGIGTGAALVSKMGFPNNRGITLYTVL--SVLSALPLF-----T 292
QY 393 PLMLAVVGAIONILSKSTKVALPDSTEMAYIPLDQOKYVKAALDVVARFGSGGA 452
DB 293 PIFAVSVIATILFLFA-----MEYGEVLAVKTVQENANANGIQGRIS----- 334
QY 453 LIQGLGVICSGISGAMPYLAIVT-----LFTIATLWVSATKL-NKLFLAQSALKEQ 504

```

Db      : 335 -VAESAGICISIGSM--FINILSAPVIMGLIIVVCGLFIATHTKLXNKSFLERDNKTEOK 391
RESULT 25
AFUB ACTPL STANDARD; PRT; 687 AA.
AC Q44123; Q44122;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update).
DE Ferric transport system permease protein afub.
GN AFUB.
OS Acinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=715;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S 4074 / Serotype 1, and K17 / Serotype 5;
RA MEDLINE=96401417; Pubmed=807793;
RX Chin N., Frey J., Chang C.F., Chang Y.F.;
RT "Identification of a locus involved in the utilization of iron by
RL Acinobacillus pleuropneumoniae."
CC FEBS Microbiol. Lett. 143:1-6(1996).
CC -1 FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR IRON. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE
CC SUBSTRATE ACROSS THE MEMBRANE.
CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1 SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U04954; AAB17217.1; -.
DR EMBL; U05042; AAB05031.1; -.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 2.
DR ProSite; PS00402; BPD_TRANS INN_MEMBER; FALSE_NEG.
KW Transports; Iron transport; Transmembrane; Inner membrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 106 126 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 378 398 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
FT TRANSMEM 484 504 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 545 565 POTENTIAL.
FT TRANSMEM 594 614 POTENTIAL.
FT TRANSMEM 620 640 POTENTIAL.
FT TRANSMEM 649 669 POTENTIAL.
SO SEQUENCE 687 AA; 75604 MW; DEC6859BC2DA0FE2 CRC64;
Query Match 4.6%; Score 122; DB 1; Length 687;
Best Local Similarity 18.7%; Pred. No. 0.97;
Matches 111; Conservative 86; Mismatches 222; Indels 176; Gaps 24;
38 TENVTVARDTK-DLLIVGAPGSGEALPFIKFMLVVPICALIFM----- 79

```

Db 122 TFAFAKLMQGDKEFIILALIS---IIILFFIIVPTVALIFSMVDEGTAPRQVRI 178
 Qy 80 ---LIYAKSNILSKQALFYAVGTPFLIFPALFPTVYPRLDVHHPPEADRLQAL-- 133
 Db 179 LGQGYIVRVISNSLFLSGFVIGVSTVFGIALAFALYTT-----RLAHRAFAFGKIFSLPI 232
 Qy 134 ---PGLLGLVAIL-----RWTFAPFV--LAELMGVMSLPMFG 170
 Db 233 VTFPFVGLGVTLMGRSGVTEFLDTYFGFRDHNMLVGFNGIALAQLAPISFMLD 292
 Qy 171 FANEITKIH-----EAKRFYALFGI-----GANIS--LLA 198
 Db 293 GA--LKSIFPSIEEASVTLRANRYQTFRYNIIFPLRLPALNSFLVFIQSLADPSNPLVL 350
 Qy 199 SGRALIVASKLRASVSEGVDPWGISLRLLMAMTVISGVLMASVMT--NKAVLDDPRRYN 257
 Db 351 GGSFVDVATQTYFYIAGSQLDYASASTIGSMILFSLIFIVQYWMIGNRSYVT----- 404
 Qy 258 PEEMQKGGKGAQPKNMKDSFLYLDSPYILLTLVLI---AVG-ICINLIEVTWKSQ 312
 Db 405 ---VSGKSYRGDQDPLPSGLKY-----TIYMLGFWVFNFALYGSIFYSFTVMGVY 456
 Qy 313 KLOYPNMNDYSEFMG---NFSFWGVSVLIMLVGQNVIRKFGMLTGALV----- 360
 Db 457 TL---TLNNYAMLFGQGLSDGAMPSLINTMIYAGIAPLTLFGLLIVIVVRKDFQKK 513
 Qy 361 -----TPVAVLTGIVFPLVYFRNOASGLVAMFGTTP 393
 Db 514 TLEFLTMLCPAVPSTVAGSVYLAENDAPMYITGIIIIISMWRDLPIKRAIAG-- 571
 Qy 394 LMLAVVGAIONILSKTKVALPDSTKEMAYI--PLDOEQKVGK-----KAALDVVAR 445
 Db 572 ---LGQDKSLDEASLSLKGSSIKTIWFIIVFLKPALLSALVYSFYAAMTVSAIV 625
 Qy 446 FGSAGGALLQGGLI---VIGSIGAMPYLAIVLLFIAT-----WIVSATKLNK 492
 Db 626 FLVYADTRVATAYILNRVEDGEYVALVYGSILIVMMAILIFPDWIGDTRIRK 680

RESULT 26
 YN86_YEAST STANDARD; PRT; 894 AA.
 ID YN86_YEAST STANDARD; PRT; 894 AA.
 AC P27514;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical 99.5 kDa protein in URK1-SM1 intergenic region.
 GN YNR013C OR N2052.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / FY1679;
 RX MEDLINE=95208356; PubMed=7900425;
 RA Verhaaselt P., Aert R., Voet M., Volckaert G.;
 RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
 the centromere on the Saccharomyces cerevisiae chromosome XIV right
 arm.";
 RT Yeast 10:1355-1361 (1994).
 RL [2]
 RP SEQUENCE OF 1-293 FROM N.A.
 RA Maurer C.T.C., Urbanus J.H.M., Planta R.J.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 784-894 FROM N.A.
 RC STRAIN=ATCC 28383 / FL100;
 RX MEDLINE=90364830; PubMed=2169608;
 RA Kern L.;
 RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
 RL Nucleic Acids Res. 18:5279-5279 (1990).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE SLC13A FAMILY OF TRANSPORTERS.
 CC PH087 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X77395; CA54581.1; -
 DR EMBL; Z71628; CA96290.1; -
 DR EMBL; Z71629; CA96292.1; -
 DR EMBL; X53998; CA37947.1; ALT_INIT.
 DR PIR; S45135; S45135.
 DR SGD; S0005296; YNR013C.
 DR GO; GO:0015114; P:phosphate transporter activity; IGI.
 DR GO; GO:0006817; P:phosphate transport; IGI.
 DR InterPro; IPR001898; Na/Sul_symp.
 DR InterPro; IPR004331; SPX.
 DR Pfam; PF00939; Na_sulph_symp; 1.
 DR Pfam; PF03105; SPX; 1.
 DR KX Hypothetical protein; Transmembrane.
 FT TRANSMEM 430 450
 FT TRANSMEM 474 494
 FT TRANSMEM 511 531
 FT TRANSMEM 557 577
 FT TRANSMEM 602 622
 FT TRANSMEM 642 662
 FT TRANSMEM 682 702
 FT TRANSMEM 706 726
 FT TRANSMEM 738 758
 FT TRANSMEM 777 797
 FT TRANSMEM 799 819
 FT TRANSMEM 824 844
 FT TRANSMEM 874 894
 SQ SEQUENCE 894 AA; 99490 MM; 21938585EB0509 CRC64;
 Query Match 4.6%; Score 121.5; DB 1; Length 894;
 Best Local Similarity 18.9%; Pred. No. 1.4;
 Matches 116; Conservative 83; Mismatches 178; Indels 237; Gaps 31;
 Qy 3 KTESKPFGRKRSFLMPHTEHKKVLPMLMFCITNTYVYPRLDVHHPPEADRLQALPFG-- 59
 Db 365 KTDKKKSKLSS-----SQLRP-----NANIT-----BSMMSGGAGIIA 400
 Qy 60 --AEAIPE-----IKFWLVVPCAIIFM----- 79
 Db 401 PSTDSLTFREMLHPKQMLQFTMGQTSLLKFLUTISCFIALTFNLPTQDSLQKNCF 460
 Qy 80 --LIYAKSNILSKQALFY-avgTPFLIFPALFPTVYPRLDVHHPPEADRLQALPFG 136
 Db 461 ALIIVASLWATERIPLFVTSIMIPILI--VFPVINDPITS--QPMSPDSSQ----- 510
 Qy 137 LILGVALIRWMTAPFVYLAELMGVMSLPMNGF--ANEITKHEAK--RFVALFGIGA 192
 Db 511 -----FISTMWSVIM-LLLGFTLAALSIXNIAKIVSLTHIASAGT 553
 Qy 193 N-----ISLLASGRAIVWASKLRASV--SEGVDPWGISRL-----MAMTVISG 235
 Db 554 NPHFILLTMFVALFPLFS---MWSVNAAPVLCYSIVP---LRLTLPRNCSTAKALLIG 606
 Qy 236 LVIMASYWIMINKVNLDPREFYNEEMQKGGKAK-----P 270
 Db 607 IALASNIGWSPSPASQNFISGIMDPSPSMAEFMIALPVCFICMAITWILLITFPF 666
 Qy 271 KMMKGSFLYLDSPYILLTLVIVAGICINLIEVTW--KSQDKLOYPMN----- 319
 Db 667 EPNVKILQHPSRDPLFKQWFTL---VCIIRI-VLMCSNIOISGIFGEMGIISIIPIV 722
 Qy 320 -----NDSEFNGNFSFMTGVSVVIMLVFGQNVIRKFGMLTGALVTP----- 362

Db 723 VFPGNGLTSDDDNNFM-----WT-----IVVLAMGTTLGKAVSSSGSLSTMAQLIAQ 772
 QY 363 -----VMULLTGVFPALVIFRNQASGLVAMGTTPLM-----LAVVVG 401
 Db 773 VEHBPFLIVLIFGLVILVMAVF--VSHTVAMMIYVLMSEIGSNLPSGHSRLIVIA 829
 QY 402 AIONILSKSTKVALFDSTKEMAVIPLDEOKYKGAIDVVAARFGKSGALIOQGLLVI 461
 Db 830 A---LLCGSA-----MGLPTSGFPNVTAISMDVGDXY-----LIVGTFIT 868
 QY 462 CGSIGAMTPYLAIV 475
 Db 869 RGVPAISLSTVAIV 882

RESULT 27
 COX1 ALUMA STANDARD; PRT; 536 AA.
 ID COX1 ALUMA STANDARD; PRT; 536 AA.
 AC P80470;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1.
 OS Allomyces macrogynus.
 OC Mitochondrion.
 OC Eukaryota; Fungi; Chytridiomycota; Blastocladiiales; Blastocladiaceae;
 OC Allomyces.
 OX NCB1_TaxID=28583;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
 RX MEDLINE=96081490; PubMed=7490780;
 RA Paquin B., Forget L., Roewer I., Lang B.F.;
 RT "Molecular phylogeny of Allomyces macrogynus: congruency between
 nuclear ribosomal RNA- and mitochondrial protein-based trees.";
 RL J. Mol. Evol. 41:657-665(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
 RX MEDLINE=96226032; PubMed=8636971;
 RA Paquin B., Lang B.F.;
 RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
 sequence from an ancestral fungus.";
 RL J. Mol. Biol. 255:688-701(1996)
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL: U41288; AAC49234.1; -
 PIR: S63651; S63651.
 HSSP: P98002; IARI.
 InterPro: IPR000883; COX1.
 Pfam: PF00115; COX1.1.
 PRINTS: PR01165; CYCOXIDASE1.

DR PROSITE: PS00077; COX1.1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 67 67 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 246 246 COPPER B (PROBABLE).
 FT METAL 250 250 COPPER B (PROBABLE).
 FT METAL 295 295 COPPER B (PROBABLE).
 FT METAL 296 296 COPPER B (PROBABLE).
 FT METAL 381 381 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 383 383 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 246 250 1'-histidyl-3'-tyrosine (By similarity).
 SQ SEQUENCE 536 AA; 59444 MW; 7DE8DF291EAE4091 CRC64;
 Query Match 4.6%; Score 121; DB 1; Length 536;
 Best Local Similarity 20.4%; Pred. No. 0.88; Indels 144; Gaps 24;
 Matches 93; Conservative 65; Mismatches 155;
 QY 22 HELKKVL---PMFLMPCITF-----NYTVLRDTKDTLLVAGPSGAELPIFIKFWL 70
 Db 57 HOLNVITTAHAFIMIFFLVMPMLGFGNYFV-----PIMGAPDMARPLNNISFWL 110
 QY 71 VVPCALIFM-----LIYAKLSNLSKQALFYAVGTFPLIFPL----- 108
 Db 111 LPSELILVGSAAFEVQAGAGTGTVPYPLSSI-----GFHSGGSVDLAIFSLHLAIGISSML 165
 QY 109 -----FFTVIYPLRD---VLHPT---PADRLQAL---PPGLLGVAII---RNVTFAA 151
 Db 166 GSINFTITLIMKRAPGNTWKHLPLFVWSLITVTLILSLVPLAGATMLLTDNNLN-TT 224
 QY 152 FVYLAELMGSVMLSLMFW-----GFANEITKIHAKRFVALFGI---GAN 193
 Db 225 FYDPAGGDPVLVQLHFWFPGHEVYIIIRGGIISQVSTSRKRIFFYLGAVVAMAS 284
 QY 194 ISLLASGRALVAMSKLRASVSEGVDPWGISRLMANTYISGLVAMASYMT---NKVY 249
 Db 285 IGLIG---FIVWSHNM-YTVGLDVTDFAYPTAATMIAVPTGKIFP---WLATLYGQNI 337
 QY 250 LIDPRFYNPEPMQK-----KKGAKPKKMMDSFLYDRSPYILL-TLV 294
 Db 338 L---YTPAYFPALGFLFLFTTGCVGWMLANSLDALDHYVVAHFHYVLSMGAVFA 393
 QY 295 IAVGICINLIEVTKWSQLKQYPMNDYSEFMGNFSPWTSVSLI---MLFVG----- 345
 Db 394 LPAGFYWIKIGITGK-----QYNEFWQGVHWMTWIFIGNVTVFFPMHFLGLNGMP 442
 QY 346 -----GNVIRKFGMLGALVTPVMVLTGI 370
 Db 443 RRPDPYDAFTQNNVSSFGSIISIVST--IVFLYGL 477

RESULT 28
 YDHP_ECOLI STANDARD; PRT; 389 AA.
 ID YDHP_ECOLI STANDARD; PRT; 389 AA.
 AC P77389;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical transport protein ydhp.
 GN YDHP OR B1657.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCB1_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).

```

RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=K12;
RA MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kashi H., Kashiimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Mochimaru K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000261; AACT4729.1; -
DR EMBL; D90809; BAA15423.1; -
DR EMBL; D90810; BAA15432.1; -
DR PIR; C64923; C64923.
DR EcGene; EG13950; ydhp.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_Transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
SQ SEQUENCE 389 AA; 40064 MW; 90A6A22CDB565B3 CRC64;

Query Match 4.6%; Score 120; DB 1; Length 389;
Best Local Similarity 22.8%; Pred. No. 0.75;
Matches 109; Conservative 62; Mismatches 167; Indels 140; Gaps 25;

QY 40 NYTVLRTKOTLIVGARGSGAE-----AIPFIKFWLVVPCALIFMLIYATLSNLSKQA 93
DB 4 NYPLP-----ALATGARGIGITTEPSWGGLFVIRAGVDVSIIPAAGMI-----SAYA 50
QY 94 LFVAVGPFLIFF-----ALFPVIYPLRDVLHPTEPADRLQALPPGL-LGVAVI 143
DB 51 VGWVGAGPLMTLLSHRARRSALIFLMAIFLGNV-----LSALAPRYMTLMSRI 101
QY 144 LRNWTFAAFYVLAELWGSVMLSLWFGPANEITYIHEAKRFYALFGIGANISLLASGRAI 203
DB 102 LITSINHGAFFGL-----GSVVA-----ASVVPKHKQASAVATWF-MGLTLANIGVPA 149
QY 204 VMAKTLASVSEGVDPNGISIRLLMAMTIYSGVLMSYWMYINCVLTDPRFYPEEKOK 263
DB 150 TWLGET-----IGWRKMFPLATAGLVISWVSLFP-----SLPK 182
QY 264 GKXGAKPRMKNKDSFLYIDRSPLYLLTLTLVIAVGICINILEVYTWKSOLKLYENNMNDYS 323

```

```

DB 183 GGAGARPE--VKKEVLAMRQVLSALLTTLVLAGAMFTL--YTYIS-----PVLOSIT 232
QY 324 EPMGNFSEFWTGVSVLMLFVG--GNVT-----RKFGULTALVTPVWVLLTGIVFF 373
DB 233 HATPVF-----VTAMLVILGVGFSIGNVGLGKLRDSVNGTLKGFLL-LMVIMLAIPL 286
QY 374 ALVIFRQASGLVAM--FGTT-----PLMIAVY-VGAIQNLISKSTKYALPDSITKEWY 424
DB 287 A---RUEFGALISVWVGAAATPVVPLQWRVRVASAPGLSSVYVIGAFNLGNAL-- 340
QY 425 IPLDQEQVKGKALIDVVAARFGKSGALLQOGL-LVYICSGIGAMTYLAVILFIYA 481
DB 341 -----GAA-----GGAIVISAGLGYSFVPMGAI VAGLALLVFMGA 377

RESULT 29
YDHP EC057 STANDARD; PRT; 389 AA.
AC Q8X625;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical transport protein ydhp.
GN YDHP OR Z2679 OR EC82366.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfale G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21165231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Itoe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005389; AAG56646.1; -
DR EMBL; AP002558; BAB35789.1; -
DR PIR; B85773; B85773.
DR PIR; F90924; F90924.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_Transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.

```

FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 160 180 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 357 377 POTENTIAL.
 SQ SEQUENCE 389 AA; 40035 MM; B538643270688816 CRC64;

Query Match 4.5%; Score 119; DB 1; Length 389;
 Best Local Similarity 22.8%; Pred. No. 0.86;
 Matches 109; Conservative 62; Mismatches 167; Indels 140; Gaps 25;

QY 40 NYTVLRDRTKTLTGAPSSGANE-----AIPFIKFWLVPCALIFMLIYAKLSNLSKQA 93
 DB 4 NYPL-----ALATGAGFIGTTERSPMGLLPVIARGVDSIPAGMLI-----SAYA 50
 QY 94 LFVAVGTPEFLIF-----ALFPIYVPLRDVLFTEFADRLQALIPGL-LGLVAVI 143
 DB 51 VGWVAGAPLMTLLSHRARSALIFLMAIFLGNV-----LSAIPDYMTLMSRI 101
 QY 144 LRNTFAAFVYLAELMGSVMSLFWGFANEITIKHEAKRYALFGIGANISLASGRAI 203
 DB 102 LTSINHGAFFGL-----GSVA-----ASVVPKHKQASAVATWF-MGLTLANIGVPA 149
 QY 204 VMAKRLASVSEGVDPWGISRLMLAMTIVSGVIVMAYSVWINKVLTDPFVPEEMOK 263
 DB 150 TWLEST-----IGRMSEFLATAGVLSWSLFF-----SLPK 182
 QY 264 GKGAAPKPMKDSFLYDRSPYILLTLTVAYGICINLIEVTKWSQKLYPNMNDYS 323
 DB 183 GGAARPE--VKKELAVMRPOVLSALTITVLAGAMPTL--YTIIS-----PYLSIT 232
 QY 324 EFMGNFSFWTGVSVLMLFVG--GNVI-----RKFGLTGALVTPVWVLTGIYEF 373
 DB 233 HATPEV-----VTAMLVIGVFGISGNVIGKADRSVNGTKGFLTL-LWVIMLAIPL 286
 QY 374 ALVIFRNASGLVM--RGTT-----PLMLAV-VGAIONILSKSTVALFSTKEMAY 424
 DB 287 A-----RNFEGAIISAVGAATPAVVPLOMRKRVASDAQSSVNTGAFNIGMNI-- 340
 QY 425 IPLDOEKVKGAIDVVAARFGKSGGALLIOGL-LVIGSIGANTPYLAVILLFI 481
 DB 341 -----GAAA-----GGAVISAGLGYSFVPMWGAIVAGLALLLVFMSA 377

RESULT 30
 NARU SALTY STANDARD; PRT; 462 AA.
 AC P37593;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrite extrusion protein 2 (Nitrite facilitator 2).
 OS NARU OR STM1576.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L72 / SGSC1412 / ATCC 700720;
 RA MEDLINE=21534948; PubMed=11677609;
 RA Mclelland W., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milius E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT L72.";
 RL Nature 413:852-856(2001).
 RN [2]
 RP SEQUENCE OF 1-130 FROM N.A.
 RC STRAIN=SL1303;
 RX MEDLINE=95011654; PubMed=7926834;
 RA Hongo E., Morimoto M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
 RA Ichimura S., Noda Y.;
 RT "The methyl viologen-resistance-encoding gene smva of Salmonella
 typhimurium";
 RL Gene 148:173-174(1994).
 CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRITE PRODUCED BY THE
 CC DYSIMILATORY REDUCTION OF NITRATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (probable).
 CC -1- MISCELLANEOUS: SINCE THE TRANSPORT IS CONTROLLED BY OXYGEN, FIVE
 CC CYS RESIDUES IN NARU MAY PLAY AN IMPORTANT ROLE IN SENSING REDOX
 CC POTENTIAL TO REGULATE THE FACILITATOR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NARX/NASA FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AE008769; AAL20494.1; -
 DR EMBL; D26057; BAA05054.1; -
 DR Stycogen; SG10245; narU
 DR InterPro; IPR004737; NO3 transporter.
 DR TIGR; TIGR00886; ZAO108; 1.
 KW Nitrate assimilation; Transport; Transmembrane; Inner membrane;
 KW Multigene family; Complete proteome.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 288 308 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 342 362 POTENTIAL.
 FT TRANSMEM 402 422 POTENTIAL.
 FT TRANSMEM 433 453 POTENTIAL.
 FT TRANSMEM 433 453 POTENTIAL.
 FT CONFLICT 61 61 N -> D (IN REF. 2).
 FT CONFLICT 126 130 TPEGV -> FLISA (IN REF. 2).
 SQ SEQUENCE 462 AA; 49751 MM; 2CDCC677B9480B53 CRC64;

Query Match 4.5%; Score 118; DB 1; Length 462;
 Best Local Similarity 20.7%; Pred. No. 1.2;
 Matches 101; Conservative 66; Mismatches 179; Indels 142; Gaps 23;

QY 69 WLVVPCALIFMLIYAKLSNLT-----LSKQALFYAVGTPEFLIPALFPYIYVLRDVLHPEEF 125
 DB 37 WISVSCILLAFCCWMLPSAVALNKNIGFNTDQFL-----TAL 78
 QY 126 ADRLOATLPGLGLVAIL--RNWTFPAFYVL-----AELMGSVMSLFW 169
 DB 79 PLSGALLRPVPSWVLPFGKRKTIVSYLLIIPCAMGRVAVGNPAPFPVFLIALLC 138
 QY 170 GFANEITIKHEAKRYALFGIGANISLASGRAIYASKLR-----ASVSEGVDPWGISLR 225
 DB 139 GFA-----GANFA--SSWGNISFFPKARQSGSALGINGLGNLGVSV 179
 QY 226 LLMAMTV-----SGVLMASTW-WIN-KNVLTPDRFPNPEEMOKK 265
 DB 180 QLIAPLVIFLPIFTFLGVQVGPDPGSLTLTNAMTWVPLAVATLAAMGANDIGSSK 239
 QY 266 KGAPKPMKDSFLYDRSPYILLTLTVIA-VGICINLIEVTKWSQKLYPNMNI-- 320

```

Db 240 ASVA-----SOLPVLKRLHMLSLLYLATFGSFIG-FAAGFAMLAKTGPDPVNIQL 291
Qy 321 -DYSEEMGNFSPFWGVSVLMLFVGNVIRKFGULTALTPV-MVLTGTGFALVIF 378
Db 292 AFGPFPIG-----ALARSAGVISPDEFSGVRYTLINLFMALFTALLFTL--- 337
Qy 379 RNQASGLVAMF-----GTPPLMAVVGAIQNIILSKSTKY--ALPDSYKE 421
Db 338 PGSGGSPSAYIVPMGLFTLAGLGSGSTFQMIANIFQI-----TLNVKILRGSSDE 390
Qy 422 MAYIPLDGEQYKGAIDVVAARFGSGGALIQGL---LVTCGS-IGAMTPYLAIVLL 477
Db 391 QA-----QREAVTDAALGFIISAIGAVGFEFIPAFGTSIALTSSPVGAKMIFILFYLA 445
Qy 478 FIATMTV 485
Db 446 CVLLTWLV 453

RESULT 31
PUTP_HAEIN STANDARD; PRT; 504 AA.
ID PUTP_HAEIN STANDARD; PRT; 504 AA.
AC P45174;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sodium/proline symporter (proline permease).
GN PUTP OR H11352.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spirigs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: CATALYZES THE SODIUM-DEPENDENT UPTAKE OF EXTRACELLULAR
CC L-PROLINE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL, U32814, AAC22999.1; -.
CC DR PIR, E64118; E64118.
CC DR TIGR, H11352; -.
CC DR InterPro, IPR001734; Na/solut_symport.
CC DR Pfam, PF00474; SSF, 1.
CC DR TIGRFAMs, TIGR00813; ssg, 1.
CC DR PROSITE, PS00456; NA_SOLUT_SYM_1; 1.
CC DR PROSITE, PS00457; NA_SOLUT_SYM_2; 1.
CC DR PROSITE, PS50283; NA_SOLUT_SYM_3; 1.
CC TRAnsport; Amino-acid transport; Transmembrane; Inner membrane;

```

```

KM Sodium transport; Symport; Complete proteome.
FT TRANSMEM 8 71 POTENTIAL.
FT TRANSMEM 51 28 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
SQ SEQUENCE 504 AA; 54898 MW; 61ED0A52A142A7F0 CRC64;

Query Match 4.5%; Score 117.5; DB 1; Length 504;
Best Local Similarity 19.1%; Pred. No. 1.4;
Matches 82; Conservative 77; Mismatches 117; Indels 153; Gaps 22;

Qy 148 TFAAFYLAELMGSVMLSMFWGRANETKHEAKFPALG--IGANISILASGRALVW 205
Db 10 TETTY-----IFGMLLGVLAAYYTNMLSD-----YILGRRIGSFVTMSAG----- 52
Qy 206 ASKLRAVSSEGVDPMGISLRLLAMTIYSGLVNASYWM-----INKRVLTDP 253
Db 53 ASDSGMLMLGL-EGAVLSGLVEGWIALIGLTIGAFYMLVAGRLRYVTELNNAITLP 111
Qy 254 RFPNPEMOKGKKAKPRMNMKDSFLYDRSPYILLTLVAVAGICI-----NLIEVT 307
Db 112 EYFH-----NRFSSHKL-----LKLVSATITLVFTIYCAGVAVGAKLFQNIPEVE 159
Qy 308 WKSQKIQYPMNDYSEPMGNFSPWGVSVLMLFVG-----NVIRKFGMLTGALV 360
Db 160 YSTAL-----WGAATAIATVFIQGFPLAVSWTDTIOATLMIFALIT 200
Qy 361 TPVAVLLT--GIVEFALVIFRNOAS--GLVAMP-GTTP-LMLAVVGAI-----QNTLS 408
Db 201 TPVAVLLSPADTAQSAVLEBAVAANKDFDLTSTTPLLSLAMGLVFGQPHILA 260
Qy 409 KSTRYALFDSYTKENAYIPLDGEQYKGAIDVVAARFGSGGALIQGLVITC--GSIG 466
Db 261 ---RFMAADSVKSL-----IKAR-----RISGMWVLCIAGIG 291
Qy 467 -----AMTPYLAIVLFIATMLVSAATKLU 493
Db 292 IGLFAIPYFANPAIAGTVNRPEQVFIELAKLIFNMIAGILHSAL-LAAVWSTLSAQL 350
Qy 494 FLAQSLKE 502
Db 351 LISSSITE 359

RESULT 32
COXI_MARPO STANDARD; PRT; 522 AA.
ID COXI_MARPO STANDARD; PRT; 522 AA.
AC P26856;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COXI OR COXI.
OS Marchantia polymorpha (Liverwort).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiales; Marchantiaceae; Marchantiales;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;

```


"Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome." J. Mol. Biol. 223:1-7(1992).

- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.

- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.

- PATHWAY: Respiratory chain; terminal step.

- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.

- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL, M68829; AAC09451.1; -.

DR PIR, S25956; S25956.

DR HSSP, P00396; 20CC.

DR InterPro, IPR000883; COX1.

DR Pfam, PF00115; COX1. 1.

DR PRINTS, P01165; CYCOXIDASRI.

DR PROSITE, PS00077; COX1. 1.

KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane; Respiratory chain; Inner membrane.

FT METAL 65 65 IRON (HEME A AXIAL LIGAND) (PROBABLE).

FT METAL 244 244 COPPER B (PROBABLE).

FT METAL 248 248 COPPER B (PROBABLE).

FT METAL 293 293 COPPER B (PROBABLE).

FT METAL 294 294 COPPER B (PROBABLE).

FT METAL 379 379 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).

FT METAL 381 381 IRON (HEME A AXIAL LIGAND) (PROBABLE).

FT CROSSLINK 244 248 1'-histidyl-3'-tyrosine (By similarity).

SO SEQUENCE 522 AA; 57551 MW; 4B88E16F03A11D01 CRC64;

Query Match 4.4%; Score 116.5; DB 1; Length 522;
Best Local Similarity 19.1%; Pred. No. 1.7;
Matches 90; Conservative 60; Mismatches 157; Indels 165; Gaps 21;

22 HELKKVL---PMPLMFCITF-----NTYVLRDTDTLIYVAPGSGAIAPIKFWL 70
55 HQLNVNLTAHAFLMIFPMVPMAMIGFGNMFV-----PLIGSPDMAFPLNNISFWL 108
71 VVPCALIFM-----LIYAKLSNLSKQALFYAVGTPFLIFPAL----- 108
109 LPSLLULLLSALVVGCGSGWTYVPLSGITS-----HSGGSVDLAFSLHLSGVSIL 163
109 ---FPVIVLPURD---VLH-----PTEPADRLQALIPGLGLVAAIL---RNWT 148
164 GSINFITITFMRRABGLTMHRLPLFVMSVLTAF---LLLSLPLVLAGAITMLLTDRFVN 220
149 FAAYVVALMELGWSLIMFGFANEIKI-----HEAKRF-----YA 186
221 -TTFPDPRGGDPILYQHLFMFGPEVYIILPFGGIISHIVSTFSKRPVGYLQMYTA 279
187 LFGIGA-----NISLASGRAIVWASKLNASYSEV-----DPWGISLRLLM 228
280 MISIGVIGFIYMAHMFVGLDVDTRAVFTAAITMIINPTGIKIFSWIATMGGSIQYKT 339
229 AMTVSGVLVMASTYWKINKVLTDPRTYNEEMQKKGAKRKNMKDSFLYLDSPYIL 288
340 PMLFAVGFIFLFTVGGITGLIVLAN-----SGVDIALHDTYVVAHFHYVL 384

QY 289 LL-TLVAVGICINLIEVTKSQLQYPMNDYSEFMGNSFWTGVSVLIMLF----- 343
DB 385 SMGAVPFLAFGFFYWKIT-----GLQYF-----ETLQIHMTFFGVNLTFPPWHF 433

QY 344 -----VGVNIRKFGWLGVALTVPVWVLTGLVFPALVYF 378
DB 434 LGLAGMRIRIPDYDAYAGWNAFSPG-----SVSVVGIFCEVVVF 476

RESULT 33

NRFE_HA6IN STANDARD; PRT; 635 AA.
AC P44944;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome c-type biogenesis protein nife.
GN NRFE OR H10936.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellales; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bule C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL, U32775; AAC22594.1; -.
DR PIR, A64162; A64162.
DR TIGR, H10936; -.
DR InterPro, IPR003567; Cyt c biog.
DR InterPro, IPR002541; CytC_asm.
DR InterPro, IPR003568; CytC_biog_Ccmf.
DR Pfam, PF01578; CytC_asm. 1.
DR PRINTS, P01410; COBIOGENESIS.
DR TIGRFAMs, TIGR00353; nrfe; 1.
KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 6 26
FT TRANSMEM 38 58
FT TRANSMEM 93 113
FT TRANSMEM 128 148
FT TRANSMEM 167 187
FT TRANSMEM 209 229
FT TRANSMEM 233 253
FT TRANSMEM 273 293
FT TRANSMEM 311 331
FT TRANSMEM 351 371

FT TRANSMEM 390 410 POTENTIAL.
 FT TRANSMEM 421 441 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 483 503 POTENTIAL.
 FT TRANSMEM 570 590 POTENTIAL.
 FT TRANSMEM 607 627 POTENTIAL.
 SO SEQUENCE 635 AA; 71273 MW; B979E54F1ED534FC CRC64;

Query Match 4.4%; Score 116; DB 1; Length 635;
 Best Local Similarity 19.9%; Pred. No. 2.1;
 Matches 103; Conservative 46; Mismatches 142; Indels 226; Gaps 21;

QY 78 FMILYAKLSNLSKQALFYAVGTPPLIFPAPAEGRDLPMLQDVGLIHPPL 137
 DB 8 FLILLATRS-----AFPLAL-VPOFGIFKKNPLI--NAAPLSTIFLATLTLGL 56
 QY 138 LG-----LVAILRNTFAFYVLAELGWSVLSLFWGFA----- 172
 DB 57 LAYSFAVDFTLEVAHAHNSQLPFPKVAATWGHGSMLEFSLSLMLAFAFNRK 116
 QY 173 NEIRKIHAKRFYALFGANISL-----ASGR----- 201
 DB 117 NDRFSAOSLSLGLICFGFAVFLFYSNPGRIFPAPAEGRDLPMLQDVGLIHPPL 176
 QY 202 -----AIWASKLRA-----SVSEGVDPWGISLRILMAMTIYSGVLMASY----- 242
 DB 177 YVGVGAVNAPNLSALITYQASQIARSKRGVLSWLFITIGIVG-AMWAIYELGW 235
 QY 243 -NMINKNVLTDPFPYNEEMOKGKGAAPRNMKDSFLYLDSPYILLTLVIAVIGIC 300
 DB 236 GGMV-----FMDPVE----- 245
 QY 301 INLEIVWKSQLKQYPMNDYSEPMGNFSGTGVSVLIMLF-VGAVNIRKFGMLGAL 359
 DB 246 -NASLMPWLLGLALHLSLM--ATERKQGVSYWTLFSLAPAFSVLGFIVRSGLTSVH 302
 QY 360 V-----TPVMVLTGIVFPAFVIFRNOASGLVAMEGTPMLAVVGAIQNLSKSTXA 414
 DB 303 AFALDNRGVLL--LIFVLTAL--AFGLPALRA----- 333
 QY 415 LFDSTKEMAVIPLDOEKVKKAIDVVAARF-GKSGGALLIQGLVLC----- 462
 DB 334 -----GSSSESANVKFOFISKSGILLNLTLTATVSTFLGTFYP 373
 QY 463 -----GSTGAMTPY-----AVILLFIITAI 482
 DB 374 MLFQAMNWSISVSGSPYENSIFPPIITAILIMIVL 410

RESULT 34
 YGED ECOLI STANDARD; PRT; 397 AA.
 ID YGED ECOLI
 AC P3196; Q46934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yged.
 GN YGED OR B2835.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Verde J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474 (1997).
 RN [2]

RP SEQUENCE OF 1-80 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94132066; PubMed=8300626;
 RA Jackowski S., Jackson P.D., Rock C.O.;
 RT "Sequence and function of the aas gene in Escherichia coli.";
 RL J. Biol. Chem. 269:2921-2928 (1994).
 RN [3]
 RP IDENTIFICATION.
 RA Rudd K.E.;
 RL Unpublished observations (NOV-1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).

CC EMBL; U29581; AB40482.1; -;
 DR EMBL; AB000367; AAC75874.1; -;
 DR EMBL; L14681; -; NOT ANNOTATED_CDS.
 DR PIR; D65066; D65066.
 DR Ecogene; EG12455; Yged.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 53 73 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 145 165 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 227 247 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 305 325 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT CONFLICT 75 80 SPARK -> TSPKAG (IN REF. 2).
 SO SEQUENCE 397 AA; 41655 MW; 0BB3B0DEA7FF534C CRC64;

Query Match 4.4%; Score 115.5; DB 1; Length 397;
 Best Local Similarity 20.7%; Pred. No. 1.5;
 Matches 86; Conservative 63; Mismatches 143; Indels 123; Gaps 21;

QY 137 LGLVAILRNMWTFAPFVLAELMGWSVLSLFWCFANEIRKIHAKRFYALFG--IGANI 194
 DB 34 LFATLALK-----AQFY-----PEWSQPIQWFFVG-----AYILFAFPVGV 73
 QY 195 SLASGRAIYVASKLR-----ASVSEGVDPWGISLRILMAMTIYSGVLMASYMINKNV 250
 DB 74 DSFAKRVMMFANGLKLGASISICFGINP-----LGYTLVG--VGAAYSPAKYGIL 124
 QY 251 TDPFPYNEEMOKGKGAAPRNMKDSFL-----YLDSPYILLTLTLVIAVIGI 299
 DB 125 -----GELTTSKLVKANGMEASTIAAILLGSVAGVLADWHVLVLAACALAYGG 176
 QY 300 CI-----NLIEVT-----WKSQKIQYPMNDYSEPMGNFSG 331
 DB 177 AVVANIYIPKLAARPGQSMLINMTRSPFNACTSLWR-----NGETRFSLVGTSLF 228
 QY 332 WTGVVSVLIMLFVGVGNIRKFGMLTGL-----VTPVMVLTGIVFPAFVIFRNOASGLV 386
 DB 229 WGAGVTIRFLIVL-----WVPVALGITDNATP--TYLMAVAIGIVGACAAAKLV 277
 QY 387 AMFGTPLMLA-VVGAIONILSKSTK-----YALFSTKEMA--YIPDOEKVKKAA 438
 DB 278 TLEIVSCWMPAGILIGVVVILFSLQHELTPAVYALMLIGWGGFVVPVNLALGERGKS 337
 QY 439 ID-----VVAARFGSGGALLQGGILVCGSIG-AMTP-----YLAVILLFIITAIL 484
 DB 338 VQAGNAIAYONTIGNSAMLMILGIYSLAVMIGIPVVEIGIGFGLFALATITAI 392

RESULT 35
 COXI_Apili STANDARD; PRT; 521 AA.
 AC P20374;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 OS Col.
 NCBI_TaxID=7469;
 OC Apidae; Apis.
 NCBI_TaxID=7469;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thorax;
 RX MEDLINE=90136028; PubMed=2559293;
 RA Crozier R.H., Crozier Y.C., Mackinlay A.G.;
 RT "The CO-I and CO-II region of honeybee mitochondrial DNA: evidence
 RL for variation in insect mitochondrial evolutionary rates.";
 [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thorax;
 RX MEDLINE=93114603; PubMed=8417993;
 RA Crozier R.H., Crozier Y.C.;
 RT "The mitochondrial genome of the honeybee Apis mellifera: complete
 RL sequence and genome organization.";
 RT Genetics 133:97-117(1993).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M23409; AAA18476.1;
 DR EMBL; L06178; AAB96799.1;
 DR PIR; A32431; A32431.
 DR HSSP; P18401; 1PRT.
 DR InterPro; IPR008883; COX1.
 DR Pfam; PF00115; COX1.1.
 DR PRINTS; PS00077; COX1.1.
 DR PROSITE; PS00077; COX1.1.
 KM Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane
 FT METAL 59
 FT METAL 238
 FT METAL 242
 FT METAL 288
 FT METAL 289
 FT METAL 374
 FT METAL 376
 FT METAL 376
 FT CROSSLINK 238
 SEQUENCE 521 AA; 59293 MW; 2149417AC981CE64 CRC64;

Query Match 4.4%; Score 115; DB 1; Length 521;
 Best Local Similarity 20.1%; Pred. No. 2;
 Matches 86; Conservative 68; Mismatches 136; Indels 138; Gaps 22;
 QY 31 FLMEFCITFNVTY--LRDRTQTLIVGAPGSGAEIPIKFWLVVPCAIIFMLIVAKLSNI 88
 DB 61 FLMEFVMPFLLIGFGNWLIPMLGSPDMAFPMMNISFWLLPSS--LEWLL--LSN- 114
 QY 89 LSKQALFY-ANGCPFLIFPFLFFVIVPLRDVLPFTFADRLQALPLGLGLVAIIIRNW 147
 DB 115 ----LFYSPGIGWTVY-----PP-----LSAYVYHS 137
 QY 148 TFAAFVYLAELMGSVMLSLMFEGFANEITKHEAKRYALFGIGANISLASGRAIYAS 207
 DB 138 SPQVDFALIFSLHNSGISIM--GSLNLMVITIMMKNSMNYD--QSLF----- 182
 QY 208 KLRASVEGVDPMGI--SLRLMAMTIVSGLVIMASYMINKVLTDPFRFYNEPMOKG 264
 DB 183 -----PMSVFYTAIIILMSLPVLGATIMLF--DRNFNTS--FFDP----- 220
 QY 265 KKGAKPKMKKDSFLYLDKRPYLLTLVLVAVGICINLIEVTWKSQLKQYPMNDYSE 324
 DB 221 MGGDPLLYOHLFEWFGHPEVYIILP---GRELISHIV-----MNSGKKE 264
 QY 325 FMGNFSFMTGVSVYILM-----LFGGVNVRKFGMLTGAIVTPYVMTLTGI--VEPAL 375
 DB 265 IFGNLSMIVYMLGIFGFIYVAHMFVTGLDVDTAVFIYSA--TMIAVPTGIKVFSWL 322
 QY 376 VIFENQ-----ASGLVAF--GTPPLMA-----VVGAIONILSKS 410
 DB 323 ATYGSKLKLNISILMSLGFIMLFTTIGLTGIMLSNSIDIILHDYVVGVHFFHYVLSMG 382
 QY 411 TKVALPDS 418
 DB 383 AVFAIISS 390
 RESULT 36
 COXI_PODAN STANDARD; PRT; 541 AA.
 AC P20681; O21208; Q35363;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COI OR COI.
 OS Podospora anserina.
 OC Mitochondrion.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Sordariomycetidae; Sordariales; Lasiosphaeriales; Podospora.
 CC NCBI_TaxID=5145;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90124722; PubMed=2558809;
 RA Cummings D.J., Michel F., McNally K.L.;
 RT "DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase
 RT subunit I mitochondrial gene from Podospora anserina: a gene with
 RL sixteen introns.";
 RL Curr. Genet. 16:381-406(1989).
 RN [2]
 RP COMPLETE GENOME.
 RC STRAIN=s;
 RX MEDLINE=90291512; PubMed=2357736;
 RA Cummings D.J., McNally K.L., Domenico J.M., Matsura E.T.;
 RT "The complete DNA sequence of the mitochondrial genome of Podospora
 RT anserina.";
 RL Curr. Genet. 17:375-402(1990).
 RN [3]
 RP SEQUENCE OF 376-541 FROM N.A.
 RC STRAIN=s;
 RX MEDLINE=89232730; PubMed=3246349;
 RA Viterny-Jamet C.;
 RT "Senescence in Podospora anserina: a possible role for nucleic acid

RT interacting proteins suggested by the sequence analysis of a
 RT mitochondrial DNA region specifically amplified in senescent
 RT cultures".
 RL Gene 74:387-398 (1988).
 RN (4)
 RP SEQUENCE OF 4-86 AND 212-242 FROM N.A.
 RX MEDLINE=88223503; PubMed=2836091;
 RA Kuack U., Osiewacz H.D., Schmidt U., Kappelhoff B., Schulte E.,
 RA Strahl U., Esser K.,
 RT "The onset of senescence is affected by DNA rearrangements of a
 RT discontinuous mitochondrial gene in Podospora anserina.";
 RL Curr. Genet. 9:373-382 (1985).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain, terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55026; CAA38777.1; -;
 DR EMBL; M28703; AAA32001.2; -;
 DR EMBL; X07119; CAA30131.1; -;
 DR EMBL; X07120; CAA30131.1; JOINED.
 DR EMBL; X07121; CAA30132.1; -;
 DR PIR; A48327; A48327.
 DR HSSP; P98002; IARI.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASRI.
 DR PROSITE; PS00077; COX1; 1.
 KM Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 68 68 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 247 247 COPPER B (PROBABLE).
 FT METAL 251 251 COPPER B (PROBABLE).
 FT METAL 296 296 COPPER B (PROBABLE).
 FT METAL 297 297 COPPER B (PROBABLE).
 FT METAL 382 382 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 384 384 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK 247 251 N -> T (IN REF. 2 AND 4).
 FT CONFLICT 23 23 S -> G (IN REF. 2 AND 4).
 FT CONFLICT 51 51 T -> I (IN REF. 3).
 FT CONFLICT 414 414
 SQ SEQUENCE 541 AA; 59737 MW; 06AF3D497D03DAF2 CRC64;
 Query Match 4.4%; Score 115; DB 1; Length 541;
 Best Local Similarity 20.7%; Pred. No. 2.1; Indels 148; Gaps 20;
 Matches 101; Conservative 60; Mismatches 179;
 QY 51 LTVGAPSGAARIPKIFWLVVPCAIIFML-----IYAKLSNLSKQ----- 92
 DB 92 LTVGSPDVAAPRLNNISFWLPPSLILVFSACIEGAGTGWTIYPPISGVSHSGPSVD 151
 QY 93 -ALP-YAVGTPELIFLALFPVIVPLRDV--LHPT--PADRLQAIL-----PGLIG 139
 DB 152 LAIFALHLSGVSSLLGANNFTITIMNMKTPSRILFKALFCFMAVIVITAVLLLSIPVLAG 211
 QY 140 LVAIL--RNMTPAFAFYVLAELMGSVMSLFWFGRANEITIKHAKR----- 184

DB 212 AITMLTDRNRN-ISFFETAGGDPILFQHLFWFGHEVYILLIPAGIITSTISAYSN 270
 QY 185 -----YALFCIGA-----NISLASGAIVYASKLRASVSEGDV-DW- 220
 DB 271 KSVFYGIGMVAAMSGIGLFIWSSHMYTVGLDVTDAFYFAATLIIAIVPGIKIFSWL 330
 QY 221 ----GISRLMAMTIVSGVLVMSYWMINKVLTDFPFYVPEEKQKKAKPKPMNKD 276
 DB 331 ATCYGSGIRLPSMFLALGFVFMFTGGLSGVILAN-----ASLDIAFHD 375
 QY 277 SFVLDKRSPTILL-----LTLVIAAGICINLEIVTKSQSL 312
 DB 376 TVYVAHHVYLSNGAVFAMSGWFYIPKRLGLNNTLSKVQWILFIGNVTFPFOH 435
 QY 313 KLYQPNM---NDYSEFNGNSFWTGVSVLIMLFVGGNVIRKFGWLTGALTVPVM-VL 367
 DB 436 FLGIQGRPRRISDYD-----AFAGWNLISF-----GSIISSVAAMLF 474
 QY 368 TGIYFPLV--IFNQASGLVAMFCTPPLMAVVAIGAILSTKRALDSTYEMAYI 425
 DB 475 LYIVYLVGESEYVAGRFPMWLPQFTDTLQ-----ALLNSYPSLEWALSSPPKPAFV 528
 QY 426 PLDQEKV 433
 DB 529 SLPLQSN 536
 RESULT 37
 COX1_THETH STANDARD; PRT; 562 AA.
 AC Q56408;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome c ba(3)
 DE subunit I) (cytochrome cba3 subunit 1).
 GN CBA3.
 OS Thermus thermophilus.
 OS Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
 OC Thermus.
 CX NCBI_TaxID=274;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HB8 / ATCC 27634;
 RX MEDLINE=95386472; PubMed=7657607;
 RA Keightley J.A., Zimmermann B.H., Mather M.W., Springer P.,
 RA Pastuszyn A., Lawrence D.M., Fee J.A.;
 RT "Molecular genetic and protein chemical characterization of the
 RT cytochrome b3 from Thermus thermophilus HB8.";
 RT J. Biol. Chem. 270:20345-20358 (1995).
 RN [2]
 RP SEQUENCE OF 336-443 FROM N.A.
 RC STRAIN=VK1;
 RX MEDLINE=97352522; PubMed=9208917;
 RA Ehrensann A., Rak A., Garber M.B., Reinbolt J., Ehrensann B.,
 RA Ehrensann C., Grunberg-Manago M., Portier C.;
 RT "Ribosomal protein S15 from Thermus thermophilus -- cloning,
 RT sequencing, overexpression of the gene and RNA-binding properties of
 RT the protein";
 RT Eur. J. Biochem. 246:291-300 (1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RC STRAIN=HB8 / ATCC 27634;
 RX MEDLINE=20237613; PubMed=10775261;
 RA Soulliane T., Buse G., Bourenkov G.P., Bartunik H.D., Huber R.,
 RA Than M.E.;
 RT "Structure and mechanism of the aberrant ba3-cytochrome c oxidase from
 RT Thermus thermophilus.";
 RL EMBO J. 19:1766-1776 (2000).
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L09121; AAB00370.1; -
 CC EMBL: Z84206; CAB06339.1; -
 CC FIR: T52481; T52481.
 CC PDB: 1EHR; 21-FEB-01.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 KM Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
 KM Hydrogen ion transport; 3D-structure.
 FT TRANSMEM 21 41
 FT TRANSMEM 74 94
 FT TRANSMEM 105 125
 FT TRANSMEM 144 164
 FT TRANSMEM 187 207
 FT TRANSMEM 227 247
 FT TRANSMEM 267 287
 FT TRANSMEM 300 320
 FT TRANSMEM 345 365
 FT TRANSMEM 385 405
 FT TRANSMEM 420 440
 FT TRANSMEM 471 491
 FT TRANSMEM 527 547
 FT METAL 72 72
 FT METAL 233 233
 FT METAL 237 237
 FT METAL 282 282
 FT METAL 283 283
 FT METAL 384 384
 FT METAL 386 386
 FT CROSSLINK 233 237
 FT TURN 16 16
 FT HELIX 17 45
 FT TURN 46 48
 FT HELIX 52 58
 FT TURN 60 61
 FT HELIX 65 77
 FT TURN 78 78
 FT HELIX 79 97
 FT TURN 98 98
 FT HELIX 103 125
 FT TURN 126 127
 FT TURN 133 134
 FT TURN 137 138
 FT STRAND 139 139
 FT HELIX 143 173
 FT TURN 175 176
 FT HELIX 181 206
 FT TURN 207 207
 FT HELIX 208 212
 FT TURN 213 214
 FT STRAND 218 219
 FT HELIX 221 232
 FT TURN 233 233
 FT HELIX 234 249
 FT TURN 250 250
 FT HELIX 251 254
 FT TURN 255 256
 FT STRAND 259 259
 FT HELIX 262 275
 FT TURN 276 279

FT HELIX 280 283
 FT TURN 285 286
 FT TURN 288 289
 FT HELIX 292 305
 FT TURN 306 306
 FT HELIX 307 326
 FT TURN 327 328
 FT TURN 332 333
 FT HELIX 334 338
 FT TURN 341 342
 FT HELIX 344 366
 FT TURN 367 367
 FT HELIX 369 371
 FT HELIX 372 375
 FT TURN 376 377
 FT STRAND 378 378
 FT TURN 379 379
 FT HELIX 380 388
 FT TURN 389 391
 FT HELIX 392 399
 FT HELIX 400 402
 FT TURN 403 403
 FT HELIX 404 409
 FT HELIX 415 444
 FT TURN 445 446
 FT STRAND 448 448
 FT HELIX 453 455
 FT TURN 457 458
 FT HELIX 461 463
 FT HELIX 464 493
 FT TURN 502 503
 FT STRAND 511 511
 FT TURN 515 516
 FT HELIX 517 524
 FT TURN 525 525
 FT HELIX 527 550
 FT TURN 551 551
 FT STRAND 556 557
 FT STRAND 559 559
 SQ SEQUENCE 562 AA; 62527 MW; 769B9E2F203617B CRC64;

Query Match 4.4%; Score 115; DB 1; Length 562;
 Best Local Similarity 21.0%; Pred. No. 2.2;
 Matches 79; Conservative 44; Mismatches 112; Indels 142; Gaps 16;

QY 64 PFKEFWLVPCATIFMLIYAKLSNLSKQALFYAVGTP-----FLIFPALEPTY----- 112
 DB 234 FIVFWLPLAVAIITY-----TILPKQAGKLVSDPMARLAFLLFLLSTPVGFNHQF 285
 QY 113 ----IYPLRDVLHPT-----EPADRLO-----A 131
 DB 286 ADPGIDPFWKMTSHSVLTLFVAVPSLMTAFYVAASLEFGKRGKGLFGWIRALPWNPA 345
 QY 132 ILPP--GLGLVA-----ILRN--WTFAPFYV-LAEIWSYMLSLMFWGF 171
 DB 346 FVAIVLGLGLGTFPGAGGIYVASFTLDYVHNTAMVPGHFLQVASLVTLLPAMSLYWL 405
 QY 172 ANEITK-IHEAKRYVALFGANISLASGAIYVASKLRAVSVEGVDPMGISRLMMAM 230
 DB 406 PNLTKGKPSIDAKR-----RLGLAVVWL-----WFL----- 430
 QY 231 TIVSGLVIMA--SYWMINKAVLTDP-RFYNEENQKKGAKPKNNMKDSFLYLDSPYI 287
 DB 431 ----GMMINAVGLHW--AGILNVPRAYIAQVDPDAIPHAAPV-----MVFNVLAGIV 476
 QY 288 LLTLTLVAVGICINLIEVTWKSQK-----LQYNNNDYSEFWGNFSFWTGVSV 338
 DB 477 LTVALLLEFTYGLFSLVLSREKKPELAEAPLPFAVVISGPEDRRLVLMDRIGFMPAVAAI 536
 QY 339 LIMLFVGANVIRKFGML 355
 DB 537 LVVLAYGPTLVQLFGHL 553

ID	COX1_ACACA	STANDARD;	PRT;	873 AA.
AC	Q37370;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Cytochrome c oxidase polypeptide I+II (Ec 1.9.3.1).			
CN	COX1/2.			
OS	Acanthamoeba castellanii (Amoeba).			
OC	Mitochondrion.			
OC	Eukaryota; Acanthamoebidae; Acanthamoeba.			
OX	NCBI_TaxId=5755;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 30010 / Neff;			
RX	MEDLINE=95147275; PubMed=7844823;			
RA	Burger G., Plante I., Loneragan K.M., Gray M.W.;			
RT	"The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba			
RT	castellanii: complete sequence, gene content and genome			
RT	organization."			
RL	J. Mol. Biol. 245:522-537(1995).			
CC	-1 FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY			
CC	CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-			
CC	3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE			
CC	CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN			
CC	CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2			
CC	AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3			
CC	AND COPPER B.			
CC	-1 CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome			
CC	c + 2 H(2)O.			
CC	-1 PATHWAY: Respiratory chain; terminal step.			
CC	-1 SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial			
CC	inner membrane.			
CC	-1 SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HEME-COPPER			
CC	RESPIRATORY OXIDASE FAMILY.			
CC	-1 SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CYTOCHROME C			
CC	OXIDASE SUBUNIT 2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/			
CC	or send an email to license@isb-sdb.ch).			
CC	-----			
DR	EMBL; U12386; AAD11820.1; -.			
DR	PIR; S53828; S53828.			
DR	HSSP; P00396; 200C.			
DR	InterPro; IPR001505; Copper_CuA.			
DR	InterPro; IPR000883; COX1.			
DR	InterPro; IPR002429; Cyt_c_ox_2.			
DR	Pfam; PF00115; COX1; 1.			
DR	Pfam; PF00116; COX2; 1.			
DR	Pfam; PF02790; COX2_TM; 1.			
DR	PRINTS; PR01165; CYCOXIDASE1.			
DR	PRINTS; PR01166; CYCOXIDASE1.			
DR	ProDom; PD000131; Copper_CuA; 1.			
DR	ProSITE; PS00077; COX1; 1.			
DR	ProSITE; PS00078; COX2; 1.			
KW	Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;			
KW	Respiratory chain; Inner membrane.			
FT	DOMAIN	1	474	COX1.
FT		475	843	
FT	TRANSMEM	29	49	POTENTIAL.
FT	TRANSMEM	70	90	POTENTIAL.
FT	TRANSMEM	91	111	POTENTIAL.
FT	TRANSMEM	114	134	POTENTIAL.
FT	TRANSMEM	159	179	POTENTIAL.
FT	TRANSMEM	197	217	POTENTIAL.
FT	TRANSMEM	248	268	POTENTIAL.

Query Match	Score 115;	DB 1;	Length 873;
Beat Local Similarity 18.1%;	Pred. No. 3.4;		
Matches 88;	Conservative 72;	Mismatches 161;	Indels 164;
Gaps 21			
51 LIVAGPAGGAELPIKFMVLVPCALIMLVAKLSNLSKQALFYAVGTPLIF-----	105		
99 LMGAPDMAFPRNLNNSIFMLPPS--LPFLCSSI-----VEFGATGTYVPLSS	148		
106 -----PALPETYVPLRDVLHPPEFADRLQALIPGSLGVALIRNWT-----FA	150		
149 IVASGGSVDLAIPTSLHLAIGISLLGAINFTTTFNKRVPGLSMHKPLPFWSVLITAF	208		
151 APLYLAELMGSVMLSL-----MTV-----GPA	172		
209 LLPFLPVLAGAITMLITDRNFNTSFDPGSGGADPILYHOLWPFQHPREVILLPAFGIV	268		
173 NEITRKHEAKRFYALFG-IGANISILASGRAIVM-----ASKLRASV	213		
269 SQIIGTFSNKSIFGVIQNVYMLSLAVLG-PIVAHMHYTVGLDVTGRAYTTAAVMKIAV	327		
214 SEGV-----DPWGISLRLMAMTYVSGVLVASYMMINKVLTDPFRPYEEMQKK	266		
328 PTGIKIFSWIMTLWOGQIVRKTPPLFVIGFLLIPTLGLTGIVLSN-----	373		
267 GAKPRMNKDSFLVLDRSFYLLLTLLVIAVGCINLIEVTWKSQKLQYPMNDYSEFM	326		
374 -AGLDIMLHDTYVVAHFHYLVMSGAV---PAFAFGFYWPMKI-----SGYTYNEMV	422		
327 GNFSFWIGVSVLT---MLFVG-----GNVIRKFGMLTGAIVTPMV	365		
423 GNVEHMLVFIVNLTFFPMHFVGLAGMPRRIPDYPDNYNNILSSF-----GSIISVS	478		
366 LLTGIVFPAVLV--IFBNQASGLVAMFGTTPMLLVVGAIQ--NIIKSKS-TKYALP	419		
479 -----IVFYYLIYLAENN-----NNTPKLIKVHSHIFAPYIVTUSKMLTTFASIN	525		
420 KEMAY 424			
526 SDSSE 530			

GN MTCYB OR COB OR CYTB.
 OS Pinguinus impennis (Great auk).
 CC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Charadriiformes; Laridae; Pinguinus.
 OX NCBI_TaxID=94623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Feather, and Skin;
 RX MEDLINE=22189004; PubMed=12200471;
 RA Moun T., Arnason U., Arnason E.;
 RT "Mitochondrial DNA sequence evolution and phylogeny of the Atlantic
 Alcididae, including the extinct great auk (*Pinguinus impennis*).";
 RL Mol. Biol. Evol. 19:1434-1439(2002).
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 complex (complex III or cytochrome b-c1 complex), which is a
 respiratory chain that generates an electrochemical potential
 coupled to ATP synthesis (By similarity).
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 or b566) is high-potential and absorbs at about 566 (By
 similarity).
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 cytochrome c1 and the Rieske protein (By similarity).
 CC -1- SIMILARITY: Belongs to the cytochrome b family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ242685; CAC80349.1; -
 DR InterPro: IPR005798; Cytb b6_C.
 DR InterPro: IPR005797; Cytb b6_N.
 DR Pfam: PF00032; cytochrome_b_C; 1.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_QQ; 1.
 DR Electon transport; Mitochondrion; Respiratory chain; Transmembrane;
 KM Heme.
 FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B566 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 380 AA; 42549 MW; A830ADC6E8FFD0C2 CRC64;
 Query Match 4.4%; Score 114.5; DB 1; Length 380;
 Beec Local Similarity 20.8%; Pred. No. 1.6; Indels 87; Gaps 14;
 Matches 64; Conservative 54; Mismatches 103;
 QY 136 GILGVALIRMTWTFATFYALAEWGSVWLSLMFMGPAEITFKIHA----- 181
 DB 117 GIILITLMA--TAFVGYLP--WG---QMSFNG-AIVITNLPSAIRYIGQTLVEMWG 167
 QY 182 -----KKPYALFGIGANISLLASGRAIWAWSKLPAVSEGVDPNGI----- 222
 DB 168 GFSVDNPTLTFRFPALHFL--LPFWIAGIALIHILFTLHESGSN--NPLGISNCDKIPFH 222
 QY 223 ---SLRLIMAMTIYSGVLWMSYWMINKNVLTDPFNVPEEMQKKKAKRPMNKDSFL 279
 DB 223 PYFSIKDILGFTIM--FLPLTTLTALFSPNLLGDPNFPANPLVTPPHIKBWFLEFAY 280
 QY 280 YLDSPSYLLTLTLVIAIGICINLIEVWKSQKLQYNNMNDSEFMNGFSFMTGVSVL 339
 DB 281 ILRGIPTNGVGVLAALAAVVLVPLPLHKS--KQRATFRPLSGFL---FWITLVANLL 334
 QY 340 IMLFVGGNVIRKFGMLTGAIVTPVWVL--LTGIVFPALVIFRNQASGLVAMFGTTPMLA 397
 DB 335 ILTWVGSQPVH-----PFIILIGLASLIVFTIL-----LLLF 367

QY 398 VVGAIGON 405
 DB 368 PIGALEN 375
 RESULT 40
 Y917_MYCTU STANDARD; PRT; 593 AA.
 AC 005909;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable transport protein RV0917.
 GN RV0917 OR MT0942 OR MTCY21C12.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ernolova M.D., Salzberg S.L.,
 RA DeJong A., Utterback T., Weidman J., Knout H., Gill J., Mikula A.,
 RA Bhalai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE BCCT (TC 2.A.15) FAMILY OF
 TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z95210; CAB08499.1; -
 DR EMBL: AE006980; AAK45189.1; -
 DR PIR: E70582; E70582.
 DR TIGR: MT0942; -
 DR Tuberculat; RV0917; -
 DR InterPro: IPR000060; BCCT_transporter.
 DR Pfam: PF02028; BCCT_1.
 DR Pfam: PF02028; BCCT_1.
 DR Pfam: PF02028; BCCT_1.
 DR TIGRPFAMs: TIGR00842; Bcct_1.
 DR PROSITE: PS01303; BCCT_1.
 KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.

THIS PAGE BLANK (USPTO)

Alignment Scores:			
Pred. No.:	7.61e-271	Length:	1230025
Score:	2617.00	Matches:	513
Percent Similarity:	99.61%	Conservative:	0
Best Local Similarity:	99.61%	Mismatches:	2
Query Match:	99.51%	Indels:	0
DB:	20	Gaps:	0
US-09-869-433-2 (1-515) x AAX91990 (1-1230025)			
QY	1	MetThrIysThrgIuGluIySProphEgIyLysLeuAArgSerPheLeuTTPProIleHis	20
DB	404370	ATGCAAAAACCGAAGAAAAACCTTTGGAAATGGCTCTTTCTTGCGCGATCAT	404429
QY	21	ThriSgIleuLysLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn	40
DB	404430	ACTCAAGGCTTAAGAAAGTTCTGCCAATGTTCTTAATGTTCTCTATATACATTAAAC	404489
QY	41	TYrThrValLeuAArgPThrLysASPThrLeuIleValGlyAlaP-ProGlySerGlyAla	60
DB	404490	TATACGGTGTACCGCATACAAAGACACTTTATGTGGAGCTCGGTCTGGGCA	404495
QY	61	GIuAlaIleP-ProPheIleLysPhePThrLeuValAlaProCysAlaIleIlePheMetLeu	80
DB	404550	GAGGCAAAACCTTTCATCAAGATTGGCTTGTGTGCTCCCTGCTATATCTTATGCT	404609
QY	81	IleTYrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTYrAlaValGlyThr	100
DB	404610	ATTATGCAAGCTTAAGTAATTTTAAAGTAGACGCGCTTATTTAAGCAGTGGAAAG	404669
QY	101	ProPheLeuIlePhePheAlaLeuPheProThrValIleTYrProLeuAArgSPValLeu	120
DB	404670	CCCTTTTAAATTTCTTTGCCCTGTTCCTGCACTGTAATTTATCCGTAACGATGTTTA	404729
QY	121	HisP-ProThrGluPheAlaSPArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu	140
DB	404730	CATCCACAGAAATTTGTCGACCGTTTACAGCGCACTCACTCCAGAGTTGCTAGACTC	404789
QY	141	ValAlaIleLeuAArgAsnTPThrPheAlaAlaPheTYrValLeuAlaGluLeuTPGly	160
DB	404790	GTTGCCATCTTAAGAAACGTGGACATTTGCTGCATTTTAATGTACTGCTGAACATATGGGGA	404849
QY	161	SerValMetLeuSerLeuMetPheTPGlyPheAlaAsnGluIleThrLysIleHisGlu	180
DB	404850	AGGCTAGCTCATCTCTTAATGTTCTGGGGAATTTGCTAAAGAAATTAACAAATTCACGAA	404909
QY	181	AlaLysAArgPheTYrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly	200
DB	404910	GCAAAACGTTTCTACCTCTTTTGGATACGAGCTTAATTAATTTCTTACTAGCTTCGCT	404969
QY	201	ArgAlaIleValIleTPAlaSerLysLeuAArgAlaSerValSerGluIleValASP-ProTP	220
DB	404970	CGTGCAATTTTGGGCTTCAAAGTTGAGAAGCTTCGCTTCGAAAGTGAAGTCACTCTTG	405029
QY	221	GlyIleSerLeuAArgLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla	240
DB	405030	GGAATTTCTTAAGCTCTTTTGAAGCTATGACTAATGTAATCTGAGCTTGTCTTAAGGCC	405089
QY	241	SerTYrTPThrIleAsnLysAsnValLeuThrASP-ProArgPheTYrAsnProGluGlu	260
DB	405090	AGTTACTGGTGAGCAATAAGAAAGTAATACCGACTCTGCTTCTTAATCCAGAAAGA	405149
QY	261	MetGlnLysGlyLysLysGlyAlaLysProLysMetAsnMetLysASP-SerPheLeuTYr	280
DB	405150	ATGCAAAAAGGGAAGAAAGGTGCTTAACCTTAATGAATGAATGAAGTAGCTTCTCTAT	405209
QY	281	LeuASP-AArgSerProTYrIleLeuLeuLeuThrLeuLeuValIleAlaTYrGlyIleCys	300
DB	405210	CTTGCTAGATCTCTTAATATCTTTATAGCTCTCTTGGTATGTGCTATGGAATTTGC	405269
QY	301	IleAsnLeuIleGluValThrTPThrLysSerGlnLeuLysLeuGlnTYrProAsnMetAsn	320

Wed Nov 26 13:51:12 2003

us-09-869-433-2.p2n.rng

Page 5

Db 405270 ATTAACCTTAATCGAAGTACTTGGAAAAAGTCAGCTGAACCTGCAATATCTTAATATGAAT 405329
Qy 321 APTTYSerGluPheMetGlyAsnPheserPheThrGlyValValSerValLeuIle 340
Db 405330 GACTATAGTAGTTCATCGGGAACCTTCTCTCGAGCTGGCGTAGTATCCGTATCTTATC 405389
Qy 341 MetLeuPheValGlyValSerValIleLeuGlyPheGlyTyrPheLeuThrGlyValValLeuVal 360
Db 405390 ATGCTATTTGTTGGTGTGTAACGCTATTCGTAAATTTGGATGGTAACTGAGCGCCCTAGTC 405449
Qy 361 ThrProValMetValLeuLeuThrGlyValValPhePheValLeuValIlePheAsn 380
Db 405450 ACTCTCTCATGCTTCTCTTCAACGAGTATCGTTTCTTCTGCTTGTATCTTTAGAAC 405509
Qy 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValValVal 400
Db 405510 CAAGCTTCTGGCGTGTGCTATGCTGTAACCTCTCTCATGCTGAGCTGCTGTC 405569
Qy 401 GValAlaIleGlnLeuIleLeuSerLeuSerThrTyrTyrAlaLeuPheAspSerThrLys 420
Db 405570 GGAGCTATACAGATATCTTTCGAAATTCACAAATAACCTCTCTTGACTCACTAA 405629
Qy 421 GlnMetAlaTyrIleProLeuAspGlnGluGlnValValValValValValValValValVal 440
Db 405630 GAAATGGCTATATCTCTCTTACCAAGCAAGCAAAAGTCAAGGTAAGCTGCTATTGAT 405689
Qy 441 ValValAlaAlaArgPheGlyLysSerGlyValAlaLeuIleGlnGlnGlyLeuValVal 460
Db 405690 GTAGTTCGCCCGCTTCCGAAATATCGAGAGAGCTTAACTCAACAGAGTTGCTGCTT 405749
Qy 461 IleCysGlySerIleGlyAlaMetThrProTyrLeuAlaValIleLeuPheIleIle 480
Db 405750 ATCTGGAGATATGTGAGCTATGACCCCTTATCTTCACTGATCTTCTTTCATCAT 405809
Qy 481 AlaIleTyrLeuValSerAlaThrLysLeuAsnLysLeuPheLeuAlaGlnSerAlaLeu 500
Db 405810 GCTATTTGCTGCTTCTTCTGCACTTAAGTAAACAATATCTTAAAGAGCTGCTGCTT 405869
Qy 501 LysGlnGlnValValAlaGlnGlnLysSerAlaProAlaSerSer 515
Db 405870 AAGAACAAAGAGTGGCTCAAGAAAGATTGAGCTCTGCTCTTCA 405914
RESULT 4
AA201425
ID AA201425 standard; DNA; 1038602 BP.
XX
AC AA201425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KM Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KM paratrachoma; inclusion conjunctivitis; genital disease; perilepeticis;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KM bartolinitis; pneumopathy; venereal lymphogranulomatosis; ss.
OS Chlamydia trachomatis.
XX
PN MO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98MO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffeats R;
XX

DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal urethritis;
CC epididymitis, cervicitis, salpingitis, perilepeticis, bartolinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
XX
Alignment Scores:
Pred. No.: 4.77e-219 Length: 1038602
Score: 2141.00 Matches: 409
Percent Similarity: 87.81% Conservative: 52
Best Local Similarity: 77.90% Mismatches: 50
Query Match: 81.41% Indels: 14
DB: 20 Gaps: 3
US-09-869-433-2 (1-515) x AA201425 (1-1038602)
Qy 1 MetThrLysThrGlnGluLysProPheGlyLysLeuArgSerPheLeuThrProIleHis 20
Db 392407 ATGACCTCAACCGCGGAAACCTTTGAAATGCGCTCTTCTTGGCGGATAC 392466
Qy 21 ThrHISgluLeuLysValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
Db 392467 ATGCATGAGCTGAGAAAGTTTGCACATGTTCCCAAGCTTTCTGATTTTCATTCAT 392526
Qy 41 TyrThrValLeuArgAspThrLysAspThrLeuIleValGlyAlaProGlySerGlyAla 60
Db 392527 TACAGATTTTGAAGATACAAAGATCTTATCTGTTACGACACCGGATCTGAGCA 392586
Qy 61 GlnAlaIleProPheIleLysPheThrLeuValProCysAlaIleLeuPheIleLeu 80
Db 392587 GAGGCCATTTCTTATTAAGTTGCTGCTAGTCCCTCTGCTGTTGTTCAATGCTG 392646
Qy 81 IleTyrAlaLysLeuSerAsnIleLeuSerLysGlnAlaLeuPheTyrAlaValGlyThr 100
Db 392647 ATCTACGCCAAGCTTAGCAATATTTTGAACAACAGGCTCTTTCTTCCAGCTGCTCA 392706
Qy 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleTyrProLeuArgAspValLeu 120
Db 392707 CCATTCGTGTAATTTCTTGCACTGTTCCCTGGGATCTACCTTCCGATATTTCTT 392766
Qy 121 HisProThrGluPheAlaAspArgGlnAlaIleLeuProGlyLeuLeuGlyLeu 140
Db 392767 CACCCAAAGCTTGGCTGATACCTTACAAATCATTCTTCCGGAATTATGGAATTC 392826
Qy 141 ValAlaIleLeuArgAsnTrpThrPheAlaAlaPheTyrValLeuAlaGluLeuTrpGly 160
Db 392827 ATTCGATGCTACCGCACTGAGCAATTTGCTGTCTACGCTTTCGAACTTTGGGGA 392886
Qy 161 SerValMetLeuSerLeuMetPheThrGlyPheAlaAsnGluIleThrLysIleHisGlu 180
Db 392887 AGCGTTATGCTCTCTTATATCTTCTGGGCTTTCGCAATGAATTAATAATTAAGCCAA 392946
Qy 181 AlaLysArgPheTyrAlaLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
Db 392947 GCTAAAGCTTTCTACGCTCTTTTGGAGTAGAGCTTAACGTAAGCTCTCTTAATTTCTGCT 393006
Qy 201 ArgAlaIleValTyrPalaSerLysLeuArgAlaSerValSerGlnGlyValAspProTrp 220

83 108.5 4.1 485 24 ABP76898
84 108.5 4.1 587 22 AAU34198
85 108.5 4.1 596 22 AAU33748
86 108.5 4.1 603 22 AAU37074
87 108.5 4.1 604 22 AAU36932
88 108 4.1 422 23 ABP30076
89 108 4.1 451 22 AAU33583
90 108 4.1 477 22 ABP47292
91 108 4.1 868 21 AAG41507
92 108 4.1 879 21 AAG41506
93 108 4.1 890 21 AAG41506
94 108 4.1 933 22 AAU36224
95 107.5 4.1 436 23 ABP29086
96 107.5 4.1 436 23 ABP29086
97 107.5 4.1 527 19 AAG8420
98 107 4.1 527 19 AAG8420
99 107 4.1 841 21 AAU1039
100 107 4.1 1043 22 AAG92893

ALIGNMENTS

RESULT 1
ID AAY90265
AAY90265 standard; Protein; 515 AA.
AC AAY90265;
XX 22-SEP-2000 (first entry)
XX C. pneumoniae ATP/ADP translocase protein sequence.
XX ATP/ADP translocase; Chlamydia infection; diagnosis; therapy.
XX Chlamydia pneumoniae.
XX MO200039157-A1.
XX 06-JUL-2000.
XX 22-DEC-1999; 99WO-CA01224.
XX 28-DEC-1998; 98US-0114060.
XX 12-MAR-1999; 99US-0123967.
XX 30-JUN-1999; 99US-0141271.
XX (CONN-) CONNAUGHT LAB LTD.
XX PA
XX Mardin AD, Oomen RP, Wang J, Dunn P;
XX WPI, 2000-452368/39.
XX N-PSDB; AAA30922.
XX Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
XX prevention and treatment of Chlamydia infection in mammals -
XX Claim 16; Fig 1; 81pp; English.
XX
XX This sequence represents the Chlamydia pneumoniae ATP/ADP translocase
XX of the invention. The protein, DNA encoding it, or a vaccine containing
XX the DNA or protein, are useful for diagnosing, preventing or treating
XX Chlamydia infection. The sequences can also be used in a method for
XX the detection of Chlamydia infection. Primers or probes derived from the
XX DNA sequence are useful in diagnostic tests for detecting Chlamydia
XX infection.
XX
XX Sequence 515 AA;

Query Match 100.0%; Score 2630; DB 21; Length 515;
Best Local Similarity 100.0%; Pred. No. 2.2e-282;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Med Nov 26 13:51:14 2003
Frames in the complo.
C. pneumoniae
Chlamydia
disease

Qy 1 MTKEEKPFGKLRSELPRIHTEILKVLPMFLFECITFNTVLRDRTLLVAGPSSGA 60
Db 1 MTKEEKPFGKLRSELPRIHTEILKVLPMFLFECITFNTVLRDRTLLVAGPSSGA 60
Qy 61 EALPFIKFWLVCALIFMLIYAKLSNLSKOLAFVAVGPFLIFLPLVDL 120
Db 61 EALPFIKFWLVCALIFMLIYAKLSNLSKOLAFVAVGPFLIFLPLVDL 120
Qy 121 HPEFADRLQALIPGGLGLVALIRMTFAFVLAELMGSVNLSTMPGFANEITKHE 180
Db 121 HPEFADRLQALIPGGLGLVALIRMTFAFVLAELMGSVNLSTMPGFANEITKHE 180
Qy 181 AKRFYALFGANISLLASGRAIWAASKIRASVSEGVDPWGISLRILMAMTIVSGVLA 240
Db 181 AKRFYALFGANISLLASGRAIWAASKIRASVSEGVDPWGISLRILMAMTIVSGVLA 240
Qy 241 SYWMINKVLTDPFRFYNPEMOKGKGAEPKNNMDSFYLRSPYITLLTLVAYGIC 300
Db 241 SYWMINKVLTDPFRFYNPEMOKGKGAEPKNNMDSFYLRSPYITLLTLVAYGIC 300
Qy 301 INLEIVTKSQKLOQYPMNDYSEFMGNSFWTVGSVILMLFVGANVIRKGMLTGALV 360
Db 301 INLEIVTKSQKLOQYPMNDYSEFMGNSFWTVGSVILMLFVGANVIRKGMLTGALV 360
Qy 361 TPVWVLTGIVPELAVIRNOASGLVAMFGTTPMLAVVGAIONLSKSTKYALFDSTK 420
Db 361 TPVWVLTGIVPELAVIRNOASGLVAMFGTTPMLAVVGAIONLSKSTKYALFDSTK 420
Qy 421 EMAYIPLEDOEKVAKAIDVVAARFGSGGALLQOGLIVCGSIGAMTPYLAIVLFTI 480
Db 421 EMAYIPLEDOEKVAKAIDVVAARFGSGGALLQOGLIVCGSIGAMTPYLAIVLFTI 480
Qy 481 AIVVSATKLNKFLAOSALKKEQVAGDSAPASS 515
Db 481 AIVVSATKLNKFLAOSALKKEQVAGDSAPASS 515

RESULT 2
ID AAY34951
AAY34951 standard; Protein; 524 AA.
XX AAY34951;
XX 13-SEP-1999 (first entry)
XX
XX Chlamydia pneumoniae transport polypeptide.
XX DE
XX Chlamydia pneumoniae transport polypeptide.
XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX vaccine; neutralising epitope.
XX OS Chlamydia pneumoniae.
XX EN WO9927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX PA (GEST) GENSET.
XX
XX Griffiths R;
XX WPI, 1999-357842/30.
XX
XX Genome sequence of Chlamydia pneumoniae
XX Page 871-872; Disclosure, 1912pp; English.
XX AAY34584-Y35879 represent the proteins encoded by all the open reading

Griffiths
(10279)

CC frames in the complete genome (see AAY3153) of *Chlamydia pneumoniae*.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY3153-135879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 524 AA;

Query Match 99.5%; Score 2617; DB 20; Length 524;

Best Local Similarity 99.6%; Pred. No. 6,2e-281;

Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTKTEKPGKLRSGFLMPLHTEHKKVLPMLFECITFNVTYLRDPTLIIVAGPSSGA 60
 DB 10 MTKTEKPGKLRSGFLMPLHTEHKKVLPMLFECITFNVTYLRDPTLIIVAGPSSGA 69
 QY 61 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 120
 DB 70 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 129
 QY 121 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLFWGPNBEITKHE 180
 DB 130 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLFWGPNBEITKHE 189
 QY 181 AKRFYALFGIGANISILASGRAIIVASKLRAVSSEGVDPWGISLRILMMATTIVSGLVMA 240
 DB 190 AKRFYALFGIGANISILASGRAIIVASKLRAVSSEGVDPWGISLRILMMATTIVSGLVMA 249
 QY 241 SYWMINKNVLTDRFVNPEEMOKGKGAAPKMMDSFYLDRSPYITLLTIVAYGIC 300
 DB 250 SYWMINKNVLTDRFVNPEEMOKGKGAAPKMMDSFYLDRSPYITLLTIVAYGIC 309
 QY 301 INLEIETWKSQOLKQYNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 360
 DB 310 INLEIETWKSQOLKQYNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 369
 QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMGTTPLMLAVVGAIONILSKSTKALPDSSTK 420
 DB 370 TPVWVLTGIVFPALVIFRNQASGLVAMGTTPLMLAVVGAIONILSKSTKALPDSSTK 429
 QY 421 EMAYVPLDQOKVKGKRAIDVVAARFGKSGGALLIOGGLVTCGSIIGAMTPYLAIVLFTI 480
 DB 430 EMAYVPLDQOKVKGKRAIDVVAARFGKSGGALLIOGGLVTCGSIIGAMTPYLAIVLFTI 489
 QY 481 AIWLVSATKLNKFLAQSALKEQEVVAOEDSAPASS 515
 DB 490 AIWLVSATKLNKFLAQSALKEQEVVAOEDSAPASS 524

RESULT 3
 ID AAY3153 standard; Protein; 529 AA.

XX AAY3153;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis transport protein.

KM Vaccines; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM bartonellosis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

OS
 XX
 PN WO9928475-A2.

PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-010707.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (BEST) GENSET.

XX Griffiths R;

DR WPI, 1999-371125/31.

PT Genome sequence of *Chlamydia trachomatis*

XX Disclousure; Page 935-936; 1755pp; English.

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of *Chlamydia trachomatis* (see AAY3153). The polypeptides can be used as
 CC vaccines against *Chlamydia trachomatis*. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. *Chlamydia*
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conjunctivitis, trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perhepatitis, bartonellosis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

SQ Sequence 529 AA;

Query Match 81.4%; Score 2141; DB 20; Length 529;

Best Local Similarity 77.9%; Pred. No. 4e-228;

Matches 409; Conservative 52; Mismatches 50; Indels 14; Gaps 3;

QY 1 MTKTEKPGKLRSGFLMPLHTEHKKVLPMLFECITFNVTYLRDPTLIIVAGPSSGA 60
 DB 2 MTKTEKPGKLRSGFLMPLHTEHKKVLPMLFECITFNVTYLRDPTLIIVAGPSSGA 61
 QY 61 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 120
 DB 62 EAIPIKFWLVVPCALIFMLIYAKLSNLSKQALFVAVGTPLIFPALFPVITYPLRDVL 121
 QY 121 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLFWGPNBEITKHE 180
 DB 122 HPTFADRLQALIPGLGLVAIIRNMTFAAFVYLAELGWSVLSLFWGPNBEITKHE 181
 QY 181 AKRFYALFGIGANISILASGRAIIVASKLRAVSSEGVDPWGISLRILMMATTIVSGLVMA 240
 DB 182 AKRFYALFGIGANISILASGRAIIVASKLRAVSSEGVDPWGISLRILMMATTIVSGLVMA 241
 QY 241 SYWMINKNVLTDRFVNPEEMOKGKGAAPKMMDSFYLDRSPYITLLTIVAYGIC 300
 DB 242 SYWMINKNVLTDRFVNPEEMOKGKGAAPKMMDSFYLDRSPYITLLTIVAYGIC 299
 QY 301 INLEIETWKSQOLKQYNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 360
 DB 300 INLEIETWKSQOLKQYNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 359
 QY 361 TPVWVLTGIVFPALVIFRNQASGLVAMGTTPLMLAVVGAIONILSKSTKALPDSSTK 420
 DB 360 TPVWVLTGIVFPALVIFRNQASGLVAMGTTPLMLAVVGAIONILSKSTKALPDSSTK 419
 QY 421 EMAYVPLDQOKVKGKRAIDVVAARFGKSGGALLIOGGLVTCGSIIGAMTPYLAIVLFTI 480
 DB 420 EMAYVPLDQOKVKGKRAIDVVAARFGKSGGALLIOGGLVTCGSIIGAMTPYLAIVLFTI 479
 QY 481 AIWLVSATKLNKFLAQSALKEQEVVAOEDSAPASS 513
 DB 480 AIWLVSATKLNKFLAQSALKEQEVVAOEDSAPASS 524

11

12

OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (825001)..(840000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (840001)..(855000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (855001)..(870000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (870001)..(885000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (885001)..(900000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (900001)..(915000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature

Alignment Scores:
 Pred. No.: 4,54e-298 Length: 1230025
 Score: 2617.00 Matches: 513
 Percent Similarity: 99.61% Conservative: 0
 Best Local Similarity: 99.61% Mismatches: 2
 Query Match: 99.51% Indels: 0
 Gaps: 0

US-09-869-433-2 (1-515) x US-09-198-452A-1 (1-1230025)

1 MetThrIysThrGluGluValProPheGlyValLeuArgSerPheLeuTrpProIleHis 20
 404370 ATGCAAAACCGAAGAAACCTTTGGAAATTTGGCTCTTCTTGCGCGATACAT 404429
 21 ThrHisGluLeuValValLeuProMetPheLeuMetPhePheCysIleThrPheAsn 40
 404430 ACTCAGCGCTTAAGAAAGCTTCTGCAATGTTCCCAATGTTCTTCTGATTAACAATTAA 404489
 41 TyrThrValLeuArgAspThrIysAspThrIleuIleValGlyValProGlySerGlyVal 60
 404490 TATACGCTGTTTCGCGATACAAAGACCTTATTTGGAGCGCTTCTGCTTCTGCTT 404549
 61 GluAlaIleProPheIleIysPheTrpLeuValValProCysAlaIleIlePheMetLeu 80
 404550 GAGGCAATCCTTTCATCAAGCTTTCGCTGTTGTCCTGCTGCTTCTTATGCTT 404609
 81 IleThrAlaIysLeuSerAsnIleLeuSerIysGlnAlaIleuPheIleValGlyThr 100
 404610 ATTATGCAAGCTAAGTATTTATTTAAGTAAACAGCCCTTATTTATGACAGTGGAGACG 404669
 101 ProPheLeuIlePhePheAlaLeuPheProThrValIleThrProLeuArgAspValLeu 120
 404670 CCCCTTTTATTTCTTTCCTGCGCTTCCGACCTGATTTATTCGCTACGCGATGTTT 404729
 121 HisProThrGluPheAlaAspArgLeuGlnAlaIleLeuProProGlyLeuLeuGlyLeu 140
 404730 CATCTCAAGAAATTTGCTGACCGTTTACAGCCATCTCACTCCAGATTCCTAAGACATC 404789
 141 ValAlaIleLeuArgAsnTrpThrPheAlaIlePheIleValLeuAlaGluLeuTrpGly 160
 404790 GTTGCATCTTAAGAACTGACATTTCTGCTATTTATTTATTTATTTATTTATTTATTT 404849
 161 SerValMetLeuSerLeuMetPheTrpGlyPheAlaAsnGluIleThrIleHisGlu 180
 404850 AGCGATGCTATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAATTCACAGAA 404909
 181 AlaIysArgPheIleValLeuPheGlyIleGlyAlaAsnIleSerLeuLeuAlaSerGly 200
 404910 GCAAGCGCTTCTACGCTTCTTTCGCTTACGAGCTTAATATTTCTTACCTCTCTGCT 404969
 201 ArgAlaIleValTrpAlaSerIysLeuAlaGlnSerValSerGluIleValAspProTrp 220

US Patent 6,559,924
 Seq. ID NO. 1
 including

Seq. ID NO. 2

102

404970 CGTGCATTTGTTGGCTTCAAGTTGAGAGCTTCGTTCTGAAAGTGTAGATCCTTGG 405029
 221 GlyIleSerLeuArgLeuLeuMetAlaMetThrIleValSerGlyLeuValLeuMetAla 240
 405030 GGAATTTCTTACGCTTTTATGAGCGATGACATTTGATCTGACCTTCTTATGAGCC 405089
 241 SerTrpTrpTrpIleAsnIysAsnValLeuThrAspProArgPheTrpAsnProGlu 260
 405090 AGTTACGCGTGAATCAATAAGAGATGACCGATCCCTGCTTCTATATCAAGAA 405149
 261 MetGluIysGlyIysIysGlyAlaIysProIysMetAsnMetIysAspSerPheLeuTrp 280
 405150 ATGCAAAAGGAGAAAGAGTGTCAAACTAAATGAATGAAGATGATCTTCTCAT 405209
 281 LeuAspArgSerProTrpIleLeuLeuLeuThrIleuValIleAlaTrpGlyIleCys 300
 405210 CTTCGTAGATTCCTTATATCTTTTATTTAGCTCTTCTGTTATTCCTATGATTTGC 405269
 301 IleAsnIleIleGluValThrTrpIysSerGlnLeuIysLeuGlnIysProAsnMetAsn 320
 405270 ATTACTTAATCAAGATGATCTTGAAGATCGCTGAATCGCAATATCTTAATATGAA 405329
 321 AspTrpSerGluPheMetGlyAsnPheSerPheTrpTrpGlyValValSerValLeuIle 340
 405330 GACTATAGAGTCAATGAGGAACTTCTCTGCTGAGCTGGCGTATGCTCTACTTATC 405389
 341 MetLeuPheValGlyIysAsnValIleArgIysPheGlyTrpLeuThrGlyAlaLeuVal 360
 405390 ATGCTATTTGTTGGTGTACCTCATCGTAATTTGATGTGATCTGAGAGCCCTAGTC 405449
 361 ThrProValMetValLeuLeuThrGlyIleValIlePhePheAlaLeuValIlePheArgAsn 380
 405450 ACTCCGTGATGCTTCTCTTAACAGATGATCGTTTCTTGGCTTCTTATCTTAAAGAC 405509
 381 GlnAlaSerGlyLeuValAlaMetPheGlyThrThrProLeuMetLeuAlaValVal 400
 405510 CAAGCTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405569
 401 GlyAlaIleGlnAsnIleLeuSerIysSerThrIysTrpAlaLeuPheAspSerThrIys 420
 405570 GAACTTACAGAAATATCTTTCGAAATCCACAAATACCTCTTCTGACTCAACTAA 405629
 421 GluMetAlaTrpIleProLeuAspGlnGlnIysValIysGlyAlaAlaIleAsp 440
 405630 GAAATGCGCTATACCTCTTTCGACCAAGAGCAAAAGTCAAGATGAGGCTGCTATGAT 405689
 441 ValValAlaAlaArgPheGlyIysSerGlyIysAlaLeuIleGlnGlnIysLeuVal 460
 405690 GATGTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405749
 461 IleCysGlySerIleGlyAlaMetThrProTrpIleuAlaValIleLeuPheIleIle 480
 405750 ATCTGTGAAGATTTGAGCTATGACCTTATCTTTCGATGATTTCTTCTTCACTAT 405809
 481 AlaIleTrpLeuValSerAlaThrIysLeuAsnIysLeuPheAlaGlnSerAlaLeu 500
 405810 GCTATTTGTTGTTTCTGCACTAATTAACAACTATTTCTTACGCGAGTCTGCTT 405869
 501 IysGluGlnGluValAlaGlnIleuPheSerAlaProAlaSerSer 515
 405870 AAGAAACAAGATGCTCAAGAAATTAAGCTCCCTGCTTCTCA 405914

RESULT 2
 US-09-198-452A-1/c
 Sequence 1, Application US/09198452A
 Patent No. 6,559,924
 GENERAL INFORMATION:
 APPLICANT: Glaxo, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragm
 therof and uses thereof, in particular for the diagnosis, f
 TITLE OF INVENTION: and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A

ALIGNMENTS

cga

162 (e)

RESULT 1
US-09-198-452A-369
Sequence 369, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffois, R.

TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 369
LENGTH: 524
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-369

Query Match 99.5%; Score 2617; DB 4; Length 524;
Best Local Similarity 99.6%; Pred. No. 6,4e-274;
Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

1  MTKTEKPGKLRSLMPHITHELKCVLPMFLMPCITFNNTVLRDTKTLVAGAPGSGA 60
DB  MTKTEKPGKLRSLMPHITHELKCVLPMFLMPCITFNNTVLRDTKTLVAGAPGSGA 60
QY  61  EAIPIKFWLVPCAIIFMLIYAKLSNISKOALFYAVGTPFLIFPAFPVITYPLRDVL 120
DB  70  EAIPIKFWLVPCAIIFMLIYAKLSNISKOALFYAVGTPFLIFPAFPVITYPLRDVL 129
QY  121 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 180
DB  130 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 189
QY  181 AKRYFALFGIGANISLASGRAIYVASKLRSVSEGVDPWGISRLIMAMTIVSGVLMA 240
DB  190 AKRYFALFGIGANISLASGRAIYVASKLRSVSEGVDPWGISRLIMAMTIVSGVLMA 249
QY  241 SYMMINKNVLTDPFVYPEEMOKGKAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGIC 300
DB  250 SYMMINKNVLTDPFVYPEEMOKGKAKPRANKKDSFLYIDRSPYLLLTLLVIAVIGIC 309
QY  301 INLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 360
DB  310 INLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGNVIRKFGMLTGALV 369
QY  361 TTPMVLITGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKSTKALDSTK 420
DB  370 TTPMVLITGIVFALVIFRNQASGLVAMFGTTPMLAVVGAIONLSKSTKALDSTK 429
QY  421 EMAYIPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISGAMTPYLAIVLLFTI 480
DB  430 EMAYIPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISGAMTPYLAIVLLFTI 489
QY  481 AIWLVSATLKLKFLAQSALKEQEVAAQEDSAPASS 515
DB  480 AIWLVSATLKLKFLAQSALKEQEVAAQEDSAPASS 524

```

polypeptides, fragments
the diagnosis, prev

TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 653
LENGTH: 551
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-653

Query Match 39.0%; Score 1025.5; DB 4; Length 551;
Best Local Similarity 39.4%; Pred. No. 6e-102;
Matches 214; Conservative 107; Mismatches 185; Indels 37; Gaps 9;

```

1  MTKTEKPGKLRSLMPHITHELKCVLPMFLMPCITFNNTVLRDTKTLVAGAPGSGA 60
DB  12  MOSSEVKEFSRLRAYLCPIYKSEFSKVPFLFLAFVGFNVCILKMKDTLVIGSDAGA 71
QY  61  EAIPIKFWLVPCAIIFMLIYAKLSNISKOALFYAVGTPFLIFPAFPVITYPLRDVL 120
DB  72  EAIPIKFWLVPCAIIFMLIYAKLSNISKOALFYAVGTPFLIFPAFPVITYPLRDVL 131
QY  121 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 180
DB  132 HPTFADRLQALIPPGILGLVAILRNMTFAAFYVLAELMGVMSLMFPGANEITKHE 191
QY  181 AKRYFALFGIGANISLASGRAIYVASKLRSVSEGVDPWGISRLIMAMTIVSGVL 238
DB  192 AKRYFALFGIGANISLASGRAIYVASKLRSVSEGVDPWGISRLIMAMTIVSGVL 250
QY  239 MASWINKNV--LTDRFVNPPEMOKGKGA-----KRNANKDSFLYIDRSPY 287
DB  251 I----WLYRIHLLITDTSIPSRVLAEGATYANLKEKKKAKARNLFLHLSRYL 306
QY  288 LLTLVIAVIGICINLIEVWKSQKLQYFNNNDYSEFMGNFSFWTGVSVLIMLFVGN 347
DB  307 LGALIVLSYNLVIHFEVWVKQVQVSISSHFVFNYSRITLLIGVSVLAIVLLTQ 366
QY  348 VIKFPGMLGALVTPMVLITGIVFALVIFRNQASGLVAMFGTTPMLAVVGAION 406
DB  367 CIRKMGVIGALVTPMVLITGIVFALVIFRNQASGLVAMFGTTPMLAVVGAION 426
QY  407 LSKSTYALFDSYKENAYIPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISG 466
DB  427 LSKSTYALFDSYKENAYIPLDQOKYKGAIDVVAARFGSGGALIOGGLVIGSGISG 486
QY  467 AMTPYLAIVLLFTIWIWVSATLKLKFLAQSALKEQEVAAQEDSAPASS 515
DB  487 ASLVNIAVLLFTIWIWVSATLKLKFLAQSALKEQEVAAQEDSAPASS 546
QY  509 DSA 511
DB  547 EMA 549

```

RESULT 3
US-09-198-452A-306
Sequence 306, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffois, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 306
LENGTH: 907
TYPE: PRT
ORGANISM: Chlamydia pneumoniae

THIS PAGE BLANK (USPTO)

ALIGNMENTS

RESULT 1

ADP, ATP carrier protein CP0408 [imported] - Chlamydia pneumoniae (strain CHL029 at E72089
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: E72089; D81580
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999.
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A12000; MUID:99206506; PMID:10192388
 A:Accession: E72089
 A:Residues: 1-515 <ARN>
 A:Molecule type: DNA
 A:Cross-references: GB:AE001619; GB:AE001363; NID:94376620; PIDN:AA018495.1; PID:9437662
 A:Experimental source: strain CHL029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: D81580
 A:Molecule type: DNA
 A:Residues: 1-515 <REA>
 A:Cross-references: GB:AE002202; GB:AE002161; NID:97189324; PIDN:AA038252.1; PID:9718933
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: adt 1; CP0408
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.5%; Score 2617; DB 2; Length 515;
 Best Local Similarity 99.6%; Pred. No. 8.7e-193;
 Matches 513; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

102-61
100% match

1 MTKTEKPGKRSFLMPHITHELKLVLPWFMCITENYTLVRLDRTKTLIVGAPGSA 60
 1 MTKTEKPGKRSFLMPHITHELKLVLPWFMCITENYTLVRLDRTKTLIVGAPGSA 60
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPPTVIYPLRDVL 120
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPPTVIYPLRDVL 120
 121 HPTFADRLQALIPGGLGVALIRNMTFAAFYVLAELMGSSVLSLFWGFANEITKHE 180
 121 HPTFADRLQALIPGGLGVALIRNMTFAAFYVLAELMGSSVLSLFWGFANEITKHE 180
 181 AKRFYALFGANISLASGRAIVMAKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 181 AKRFYALFGANISLASGRAIVMAKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVGC 300
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVGC 300
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWGVSVLIMLVGNAVIRKFGMLTGALV 360
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWGVSVLIMLVGNAVIRKFGMLTGALV 360
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 421 EMAYIPLDQEKVKGAIDVVAARFGKSGGALLIQOGLVIVCGSIGAMTPYLAIVTLFTI 480
 421 EMAYIPLDQEKVKGAIDVVAARFGKSGGALLIQOGLVIVCGSIGAMTPYLAIVTLFTI 480
 481 AIWVSATKLNKFLAQSALKEQVADSDAPASS 515
 481 AIWVSATKLNKFLAQSALKEQVADSDAPASS 515

RESULT 2

ADP/ATP translocase [imported] - Chlamydia pneumoniae (strain J138) E86534
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: E86534
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
 A:Reference number: A86491; MUID:20330349; PMID:10871362
 A:Accession: E86534
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-515 <STO>
 A:Cross-references: GB:BA000008; NID:98978723; PIDN:BA98559.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: adt 1
 C:Superfamily: rickettsial-type ATP/ADP translocase

Query Match 99.2%; Score 2610; DB 2; Length 515;
 Best Local Similarity 99.4%; Pred. No. 3e-192;
 Matches 512; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 MTKTEKPGKRSFLMPHITHELKLVLPWFMCITENYTLVRLDRTKTLIVGAPGSA 60
 1 MTKTEKPGKRSFLMPHITHELKLVLPWFMCITENYTLVRLDRTKTLIVGAPGSA 60
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPPTVIYPLRDVL 120
 61 EAIPFIKFWLVPCALIFMLIVAKLSNLSKQALFYAVGTPFLIFPALPPTVIYPLRDVL 120
 121 HPTFADRLQALIPGGLGVALIRNMTFAAFYVLAELMGSSVLSLFWGFANEITKHE 180
 121 HPTFADRLQALIPGGLGVALIRNMTFAAFYVLAELMGSSVLSLFWGFANEITKHE 180
 181 AKRFYALFGANISLASGRAIVMAKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 181 AKRFYALFGANISLASGRAIVMAKLRASVSEGVDPGICSLRLMAMTIVSGVLMA 240
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVGC 300
 241 SYMMINKNVLTDPRFYNPEEMQKKGAKPRMKNKDSFLYLDSPYLLTLVIAVGC 300
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWGVSVLIMLVGNAVIRKFGMLTGALV 360
 301 INLEVTWKSQKLQYPMNDYSEFMGNFSFWGVSVLIMLVGNAVIRKFGMLTGALV 360
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 361 TPWVLLTGIVFPFALVIFRNQASGLVAMFGTTPMLAVVGAIONILSKSTKVALFDSTK 420
 421 EMAYIPLDQEKVKGAIDVVAARFGKSGGALLIQOGLVIVCGSIGAMTPYLAIVTLFTI 480
 421 EMAYIPLDQEKVKGAIDVVAARFGKSGGALLIQOGLVIVCGSIGAMTPYLAIVTLFTI 480
 481 AIWVSATKLNKFLAQSALKEQVADSDAPASS 515
 481 AIWVSATKLNKFLAQSALKEQVADSDAPASS 515

RESULT 3

probable adp/ATP translocase - Chlamydia trachomatis (serotype D, strain UN3/Cx) C71561
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 20-Aug-1999
 C:Accession: C71561
 R:Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche Science 282, 754-759, 1999
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: C71561
 A:Status: preliminary

THIS PAGE BLANK (USPTO)